

Estimation du biais de la copule conditionnelle pour un test d'indépendance conditionnelle

par

Steven Fortier

mémoire présenté au département de mathématiques en vue de l'obtention
du grade de maître ès sciences (M.Sc.)

FACULTÉ DES SCIENCES
UNIVERSITÉ DE SHERBROOKE

Sherbrooke, Québec, Canada, juillet 2015

Le 17 juillet 2015

*Le jury a accepté le mémoire de Monsieur Steven Fortier
dans sa version finale.*

Membres du jury

Professeur Taoufik Bouezmarni
Directeur de recherche
Département de mathématiques

Professeur Sévérien Nkurunziza
Évaluateur interne
Département de mathématiques

Professeur Éric Marchand
Président-rapporteur
Département de mathématiques

SOMMAIRE

Ce mémoire traite principalement de deux thèmes : l'estimation de la dérivée de la fonction de répartition conditionnelle et la construction d'un test d'indépendance basé sur la copule conditionnelle utilisant notre estimateur.

Dans le premier chapitre, on fait une brève introduction aux méthodes d'estimation non paramétriques de la fonction de répartition conditionnelle et de la copule conditionnelle. Par la suite, on présente un estimateur de la dérivée de la fonction de répartition conditionnelle nous permettant d'estimer le biais du processus de copule conditionnelle présenté dans le Chapitre 1. Le troisième chapitre traite du test d'indépendance basé sur la copule conditionnelle utilisant les résultats du deuxième chapitre.

REMERCIEMENTS

J'aimerais tout d'abord remercier mon directeur de recherche, Taoufik Bouezmarni pour son temps et sa patience durant ma maîtrise. Il m'a toujours poussé à donner le meilleur de moi-même et a contribué grandement à ma formation comme statisticien. Je tiens aussi à remercier mon collègue Félix Camirand pour son aide et son temps qui m'ont permis d'avoir une meilleure compréhension de certains concepts avancés en statistiques. J'aimerais remercier tous mes amis et ma famille qui m'ont remonté le moral dans les moments plus déprimants. Pour finir, j'ai eu une belle expérience à l'Université de Sherbrooke grâce aux professeurs de mathématiques.

Steven Fortier
Sherbrooke, juin 2015

Table des matières

SOMMAIRE	iii
REMERCIEMENTS	iv
TABLE DES MATIÈRES	v
LISTE DES TABLEAUX	vii
LISTE DES FIGURES	viii
INTRODUCTION	1
Chapitre 1 — Estimation de la distribution et copule conditionnelle	3
1.1 Estimateur de la distribution conditionnelle	3
1.1.1 Résultats sur les noyaux	4
1.1.2 Sélection du paramètre de lissage	14
1.1.3 Normalité asymptotique et convergence uniforme de F_{xh}	20

1.2	Estimation de la copule	21
1.2.1	Cas non-conditionnel	21
1.2.2	Cas conditionnel	24
1.2.3	Sélection du paramètre de lissage	29
Chapitre 2 — Estimation de la dérivée de la distribution conditionnelle		33
2.1	Estimateur à noyau de la dérivée de la fonction de distribution conditionnelle	33
2.2	Convergence en loi	43
2.3	Convergence uniforme en probabilité de \hat{F}_{xh}	46
Chapitre 3 — Tests d'indépendance conditionnelle et simulations		51
3.1	Construction du test	51
3.2	Réplication de \mathbb{G}_{xh}^0	53
3.3	Validité du test et simulations	54
CONCLUSION		57
ANNEXE		58
BIBLIOGRAPHIE		69

Liste des tableaux

3.1	Niveau du test d'indépendance conditionnelle calculé avec $x = 0$ pour le DGP1 et $x = 0$ pour le DGP2.	55
3.2	Puissance du test d'indépendance conditionnelle calculée avec $x = 3$ pour le DGP1 et $x = 0.5$ pour le DGP2.	56

Table des figures

1.1	Noyau d'Epanechnikov et noyau triweight	5
1.2	Graphique des observations $\{(x_i, y_i)\}_{i=1}^{250}$ utilisées pour l'estimation de la distribution conditionnelle selon deux paramètres de lissage. Les observations notées par des "+" sont utilisées pour l'estimation de F_{xh}	9
1.3	Biais intégré de l'estimateur de la distribution conditionnelle de l'Exemple 1.1.1 pour une valeur fixe de x avec les deux types de poids : NW et LL.	12
1.4	Biais intégré de l'estimateur de la distribution conditionnelle de l'Exemple 1.1.1 pour une valeur fixe de h avec les deux types de poids : NW et LL.	13
1.5	Variance intégrée de l'estimateur de la distribution conditionnelle de l'Exemple 1.1.1 pour une valeur fixe de x	14
1.6	AMISE de l'estimateur de la distribution conditionnelle de l'Exemple 1.1.1 pour une valeur fixe de x	16
1.7	Fonctions de validation croisée de l'Exemple 1.1.1 pour $x = 0.5$. Les graphiques (a) et (b) utilisent les poids de Nadaraya-Watson et les noyaux d'Epanechnikov et triweight respectivement, alors que (c) et (d) utilisent les poids locaux linéaires avec les noyaux d'Epanechnikov et triweight respectivement.	18

1.8	Estimation de la distribution conditionnelle de l'Exemple 1.1.1 avec le noyau triweight pour $x = 0.5$ basée sur un échantillon i.i.d. de taille $n = 250$ utilisant les paramètres de lissage obtenues de la minimisation de la fonction de validation croisée.	19
1.9	Estimation de F_x de l'Exemple 1.1.1 avec le noyau triweight pour $x = 0.9$ basée sur un échantillon i.i.d. de taille $n = 250$ utilisant les paramètres de lissage obtenues de la minimisation de la fonction de validation croisée. .	20
1.10	Fonctions de validation croisée de l'Exemple 1.1.2 pour $x = 1.7$. Les graphiques (a) et (b) utilisent les poids de Nadaraya-Watson et les noyaux d'Epanechnikov et triweight respectivement, alors que (c) et (d) utilisent les poids locaux linéaires avec les noyaux d'Epanechnikov et triweight respectivement.	21
1.11	Estimation de la distribution conditionnelle de l'Exemple 1.1.2 avec le noyau triweight pour $x = 1.7$. basée sur un échantillon i.i.d. de taille $n = 250$ utilisant les paramètres de lissage obtenues de la minimisation de la fonction de validation croisée.	22
1.12	Graphique des observations $\{(y_{1i}, y_{2i})\}_{i=1}^n$ de l'Exemple 1.2.1.	26
1.13	Graphique des observations $\{(y_{1i}, y_{2i})\}_{i=1}^n$ séparées en deux groupes de l'Exemple 1.2.1 : (a) $ x_i \leq 0.2$ et (b) $ x_i \geq 0.8$	27
2.1	Approximation de l'erreur quadratique intégrée de l'estimateur de la dérivée de la distribution conditionnelle pour plusieurs tailles d'échantillon pour le modèle donné dans l'Exemple 2.1. L'expérience est faite pour trois valeurs de la covariable $X(x = 0.5(\text{bleu}), x = 0.6(\text{rouge})$ et $x=0.7(\text{vert})$. .	42

INTRODUCTION

La modélisation de distribution multidimensionnelle peut être une tâche difficile. La structure de dépendance entre les variables n'est pas toujours facile à caractériser et dans beaucoup de cas, on ne peut pas supposer la normalité de la distribution. Soit (Y_1, Y_2) de fonction de répartition conjointe H et de marges F_1 et F_2 . Sklar (voir [Skl59]) a démontré qu'il existe une copule $C : [0, 1]^2 \rightarrow [0, 1]$ unique si F_1 et F_2 sont continues telle que

$$H(y_1, y_2) = C(F_1(y_1), F_2(y_2)).$$

On peut donc modéliser d'une part la distribution de chacune des variables et d'une autre part, leur structure de dépendance avec la copule. On peut aussi retrouver la relation suivante :

$$C(u, v) = H(F_1^{-1}(u), F_2^{-1}(v)),$$

où F_1^{-1} et F_2^{-1} sont les fonctions de quantile. Si Y_1 et Y_2 sont indépendantes, alors la copule est donnée par

$$C_0(u, v) = uv.$$

Si on possède un échantillon $\{(Y_{1i}, Y_{2i})\}_{i=1}^n$ i.i.d, alors l'estimateur non paramétrique le plus commun est la copule empirique donnée par

$$C_n(u, v) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(Y_{1i} \leq F_{1n}^{-1}(u), Y_{2i} \leq F_{2n}^{-1}(v)),$$

où F_{jn}^{-1} est l'inverse généralisé de la fonction de répartition empirique de Y_j pour $j = 1, 2$. [Deh81] a développé un test d'indépendance basé sur le processus de copule

$$\mathbb{C}_n = \sqrt{n}\{C_n - C_0\}$$

en utilisant le statistique de Cramér-von Mises.

La zone de rejet du test correspond normalement au quantile d'ordre $1 - \alpha$ de la loi asymptotique de la statistique sous H_0 . Cette loi asymptotique est difficile à obtenir, car elle contient des quantités inconnues. En pratique, on doit l'approximer en utilisant des répliques valides du processus de copule.

Dans un cadre conditionnel, on pourrait être intéressé à tester l'indépendance entre (Y_1, Y_2) pour différentes valeurs d'une covariable X . Basé sur un échantillon i.i.d, [GVO10] ont développé deux estimateurs non paramétriques nécessitant l'utilisation de paramètres de lissage. Ils ont par la suite analysé le comportement asymptotique de deux processus semblables à \mathbb{C}_n . Dans ce mémoire, on base le test d'indépendance sur le premier estimateur de la copule conditionnelle de [VOG11]. Malheureusement, l'une des quantités inconnues du processus est le biais asymptotique de cet estimateur. Celui-ci est exprimé en terme de la dérivée par rapport à x de la fonction de répartition conditionnelle qui est également inconnue. Le premier chapitre de ce mémoire est consacré à l'estimation non paramétrique de la fonction de répartition conditionnelle ainsi que de la copule conditionnelle. Dans le Chapitre 2, on propose un estimateur de la dérivée par rapport à x de F_x et on démontre certaines propriétés nous permettant d'estimer le biais de l'estimateur de la copule conditionnelle. Ensuite, dans le Chapitre 3, on décrit plus en détail le fonctionnement du test d'indépendance et on fait certaines simulations pour valider le test.

Chapitre 1

Estimation de la distribution et copule conditionnelle

1.1 Estimateur de la distribution conditionnelle

Soit X , une variable aléatoire de fonction de densité μ . Souvent, on s'intéresse à estimer la fonction de répartition de Y pour une valeur fixe $X = x$ à partir d'un échantillon de n observations $(Y_i, X_i)_{i=1}^n$ indépendantes et identiquement distribuées de la distribution H . La distribution empirique est l'estimateur le plus populaire de la fonction de répartition non conditionnelle de Y . Elle est notée F_n et est définie par

$$F_n(y) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(Y_i \leq y).$$

Elle attribue le même poids à chaque observation. Pour définir un estimateur de la fonction de répartition conditionnelle, on doit d'abord présenter la notion de noyau d'ordre deux. La prochaine sous-section établit des résultats portant sur les noyaux.

1.1.1 Résultats sur les noyaux

Définition 1.1.1 (Noyau). Soit $K : \mathbb{R} \longrightarrow \mathbb{R}^+$, on dit que K est un noyau d'ordre k si :

1. $\int K(u)du = 1$,
2. $\int u^j K(u)du = 0 \quad \forall j = 1, \dots, k-1$,
3. $\int u^k K(u)du \neq 0$.

Dans ce mémoire, on considère les noyaux d'ordre deux et définis sur $[-1, 1]$ pour simplifier les développements. Les méthodes d'estimation à noyau sont très utilisées dans le domaine de l'estimation non paramétrique.

Exemple 1.1. *Les noyaux d'Epanechnikov et triweight sont deux types de noyaux d'ordre deux que nous utilisons dans ce mémoire. Ils sont définies de la façon suivante :*

$$E(u) = 3/4(1 - u^2)\mathbb{1}(|u| \leq 1) \quad \text{et} \quad T(u) = \frac{35}{32}(1 - u^2)^3\mathbb{1}(|u| \leq 1).$$

La Figure 1.1 illustre ces deux noyaux.

La prochaine proposition énonce certains résultats de convergence de sommes de variables aléatoires i.i.d. Pour ce faire, on rappelle la loi faible des grands nombres qui va nous être utile dans tous les résultats.

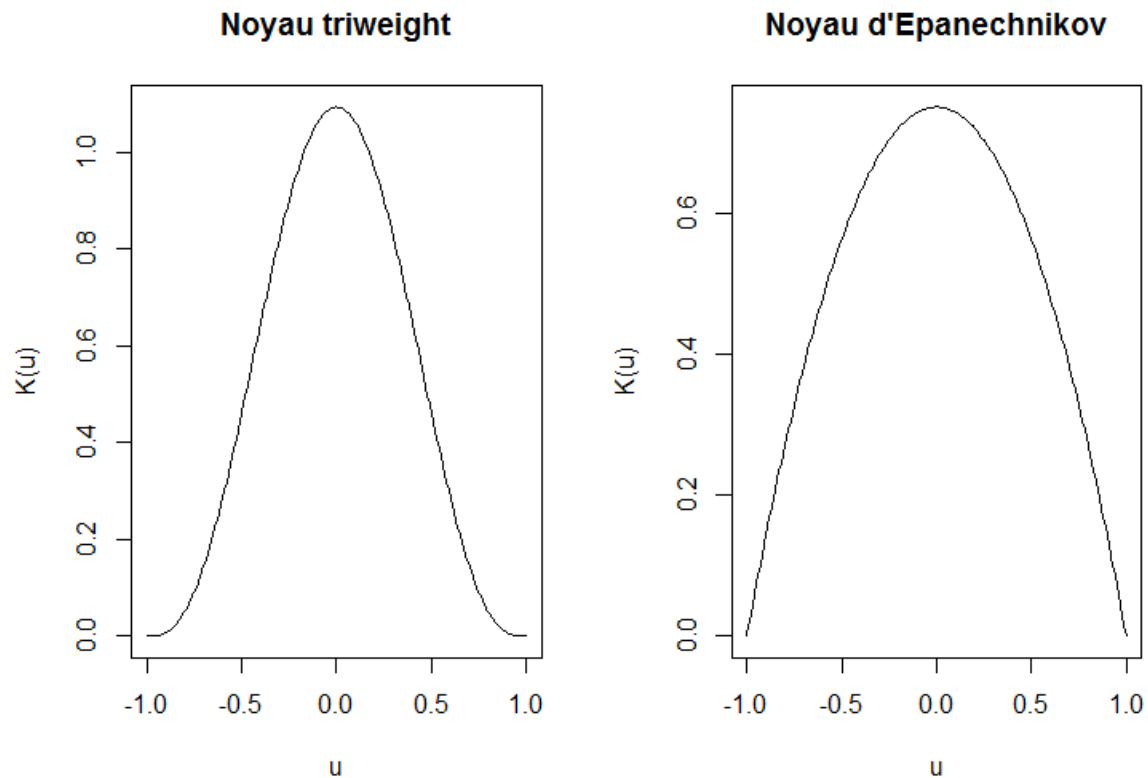


FIGURE 1.1 – Noyau d'Epanechnikov et noyau triweight

Loi faible des grands nombres (voir [vdV00])

Soit $\{Z_i\}_{i=1}^n$ une suite d'observations i.i.d tel que $E(Z_i) = \nu$ et $\text{Var}(Z_i) \leq \infty$. Si $\bar{Z} = \frac{1}{n} \sum_{i=1}^n Z_i$, alors \bar{Z} converge en probabilité vers ν et on note :

$$\bar{Z} \xrightarrow{P} \nu.$$

Dans les sections suivantes, on utilisera des développements de Taylor de μ , on a donc besoin de certaines notations et hypothèses.

On pose $\mu'(x) = \frac{d}{dx}\mu(x)$, $\mu''(x) = \frac{d}{dx}\mu'(x)$, $\kappa = \int_{-1}^1 v^2 K(v)dv$ et $R(\phi) = \int_{-1}^1 \phi^2(v)dv$ où ϕ est une fonction de $\mathbb{R} \rightarrow [0, 1]$.

Soit μ , la densité de X telle que

1. $z \rightarrow \mu'(z)$ est uniformément continue en x ,
2. $z \rightarrow \mu''(z)$ est uniformément continue en x .

Proposition 1.1. *Soit $\{X_i\}_{i=1}^n$ un échantillon i.i.d. tiré de X ayant pour densité μ deux fois différentiable et K , un noyau d'ordre deux défini sur $[-1, 1]$. Soit h_n tel que $h_n \rightarrow 0$ et $nh_n^3 \rightarrow \infty$. Alors, pour x à l'intérieur du support de X , on a :*

1. $\frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right) \xrightarrow{P} \mu(x),$
2. $\frac{1}{nh_n^3} \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right) (X_i - x) \xrightarrow{P} \kappa \mu'(x),$
3. $\frac{1}{nh_n^3} \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right) (X_i - x)^2 \xrightarrow{P} \kappa \mu(x),$
4. $\frac{1}{nh_n} \sum_{i=1}^n \left\{ K\left(\frac{X_i - x}{h_n}\right) \right\}^2 \xrightarrow{P} R(K) \mu(x).$

□

On retrouve rarement les démonstrations de ces résultats en détail dans la littérature. Nous avons jugés qu'il était pertinent de les refaire avec nos propres techniques et avec plus de détails. Les développements se retrouvent dans l'annexe et seront utiles pour le chapitre trois.

On regarde maintenant $\hat{\mu}$, un estimateur non paramétrique de μ donné par

$$\hat{\mu}(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right), \quad (1.1)$$

où h_n est un paramètre de lissage contrôlant le nombre d'observations à considérer dans l'estimation. De plus, h_n joue un rôle très important dans l'ampleur du biais et de la

variance de $\hat{\mu}$. On retrouve plus d'information dans les ouvrages suivants au sujet de l'estimation des fonctions de densité : [Sil86], [MP87], [Ize91], [Loa99] et [She04].

Le lemme suivant présente certaines propriétés de $\hat{\mu}$ démontrées dans [JW94].

Lemme 1.1.1. Soit μ la densité deux fois différentiable. On suppose que $h_n \rightarrow 0$ tel que $nh_n \rightarrow \infty$ quand $n \rightarrow \infty$, alors

1. $E(\hat{\mu}(x)) = \mu(x) + \frac{1}{2}\kappa\mu''(x)h_n^2 + o(h_n^2),$
2. $\text{Var}(\hat{\mu}(x)) = \frac{1}{nh_n}R(K)\mu(x) + o((nh_n)^{-1}),$
3. $\sup_{x \in \mathbb{R}} |\hat{\mu}(x) - \mu(x)| \xrightarrow{P} 0$ quand $n \rightarrow \infty$.

On retrouve les démonstrations dans [RL06].

En présence d'un échantillon $\{(X_i, Y_i)\}_{i=1}^n \sim H$, on estime la distribution conditionnelle de $Y|X = x$ en donnant plus d'importance aux observations (Y_i, X_i) dont la valeur X_i est proche de x . Puisque la fonction de répartition conditionnelle peut s'écrire sous la forme d'espérance conditionnelle :

$$F_x(y) = P(Y \leq y|X = x) = E(\mathbf{1}(Y \leq y)|X = x),$$

on peut utiliser les travaux de [Nad64] et [Wat64] pour retrouver l'estimateur suivant :

$$F_{xh}(y) = \sum_{i=1}^n \mathbf{1}(Y_i \leq y)w_{ni}(x, h_n), \quad y \in \mathbb{R} \tag{1.2}$$

où $w_{ni}(x, h_n)$ sont des poids étant plus grands lorsque X_i est près de x . De plus, on remarque que prendre $w_{ni}(x, h_n) = \frac{1}{n}$ revient à utiliser la distribution empirique attribuant un poids constant. Les poids sont données en fonction d'un noyau et d'un paramètre de lissage comme pour l'estimation de la densité. Les ouvrages suivants contiennent de l'information supplémentaire sur l'estimation de la fonction de répartition : [Sar93], [BHP98], [RL06], [LLR13] et [Dut15].

Dans ce mémoire, on considère seulement deux types de poids :

(NW) Les poids de Nadaraya-Watson (voir [Nad64] ou [Wat64]) sont donnés par

$$w_{ni}(x, h_n) = \frac{K\left(\frac{X_i - x}{h_n}\right)}{\sum_{j=1}^n K\left(\frac{X_j - x}{h_n}\right)}.$$

(LL) Les poids locaux linéaires (voir [GF96]) sont définies par

$$w_{ni}(x, h_n) = \frac{\frac{1}{nh_n} K\left(\frac{X_i - x}{h_n}\right) \left(S_{n,2} - \frac{X_i - x}{h_n} S_{n,1}\right)}{S_{n,0} S_{n,2} - S_{n,1}^2},$$

où

$$S_{n,j} = \frac{1}{nh_n} \sum_{i=1}^n \left(\frac{X_i - x}{h_n}\right)^j K\left(\frac{X_i - x}{h_n}\right), j = 0, 1, 2.$$

[Ize91] affirment que le choix du noyau n'influence pas de beaucoup la précision de l'estimateur. Par contre, le choix du paramètre de lissage est crucial pour la performance de $\hat{\mu}$. On suppose que $h_n \rightarrow 0$ pour mieux considérer les observations proches de $X = x$. Cependant, on veut également avoir des observations pour estimer F_x , alors on demande que $nh_n \rightarrow \infty$. La Figure 1.2 représente les observations qu'on garderait autour de $x = 0$ avec deux valeurs différentes de h_n .

La méthode de sélection du paramètre de lissage utilisée dans ce mémoire est celle proposée par [LLR13]. C'est une version améliorée de validation croisée basée sur les résultats de [Sar93]. Pour l'instant, on s'intéresse à savoir comment h_n influence le biais et la variance de F_{xh} .

Posons maintenant $\dot{F}_x(y) = \frac{\partial}{\partial x} F_x(y)$ et $\ddot{F}_x = \frac{\partial^2}{\partial x^2} F_x(y)$. On suppose que

(F1) $z \rightarrow \dot{F}_z$ est uniformément continue dans un voisinage de x .

(F2) $z \rightarrow \ddot{F}_z$ est uniformément continue dans un voisinage de x .

Avec cette notation et les poids de Nadaraya-Watson, il est possible de réécrire F_{xh} de la façon suivante :

$$F_{xh}(y) = \frac{1}{nh_n \hat{\mu}(x)} \sum_{i=1}^n \mathbb{1}(Y_i \leq y) K\left(\frac{X_i - x}{h_n}\right).$$

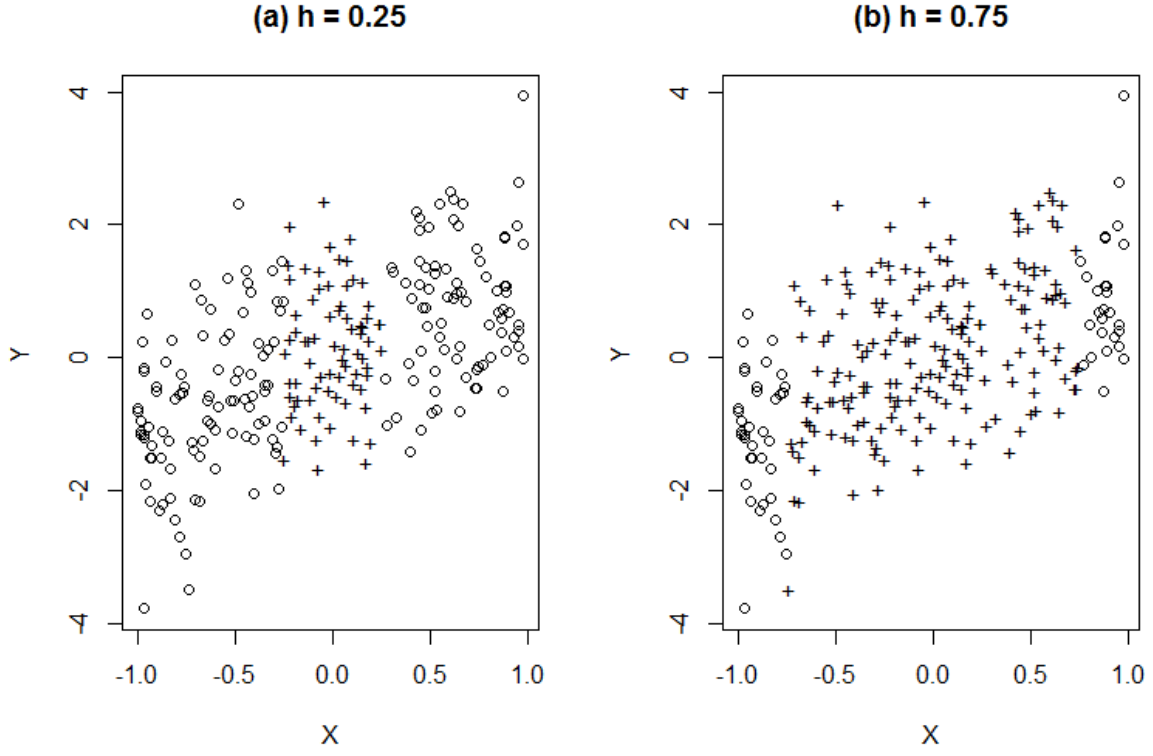


FIGURE 1.2 – Graphique des observations $\{(x_i, y_i)\}_{i=1}^{250}$ utilisées pour l'estimation de la distribution conditionnelle selon deux paramètres de lissage. Les observations notées par des "+" sont utilisées pour l'estimation de F_{xh} .

On utilise l'expression suivante pour faire des développements et ainsi contourner la double structure stochastique présente dans F_{xh} . Soit $\tilde{M}_x(y) = (F_{xh}(y) - F_x(y))\hat{\mu}(x)$. On obtient alors que

$$F_{xh}(y) - F_x(y) = \frac{\tilde{M}_x(y)}{\hat{\mu}(x)}. \quad (1.3)$$

Proposition 1.2. *Soit F_x tel que (F1) et (F2) sont vérifiées et prenons le cas où on utilise les poids de Nadaraya-Watson. Soit h_n tel que $h_n \rightarrow 0$ et $nh_n \rightarrow \infty$ et $\mu(x) > 0 \quad \forall x \in \mathbb{R}$, alors*

$$F_{xh}(y) = F_x(y) + \frac{\tilde{M}_x(y)}{\mu(x)} + O_p(h_n^2 + (nh_n)^{-\frac{1}{2}}).$$

Démonstration. [RL06] considère le cas où X est multidimensionnelle dans sa démonstration et elle peut être difficile à lire. Ils omettent aussi certains détails à la fin de leur démonstration. On refait ici la démonstration dans le cas où X est unidimensionnelle.

On peut réécrire l'équation (1.3) de la façon suivante :

$$F_{xh}(y) = F_x(y) + \frac{\tilde{M}_x(y)}{\mu(x)} + \left(\frac{1}{\hat{\mu}(x)} - \frac{1}{\mu(x)} \right) \tilde{M}_x(y).$$

On sait déjà que $\hat{\mu}(x) \xrightarrow{P} \mu(x)$. On peut appliquer le lemme de Slutsky (voir [vdV00]) pour obtenir :

$$\frac{1}{\hat{\mu}(x)} \xrightarrow{P} \frac{1}{\mu(x)}.$$

De plus, on peut montrer que l'espérance de $\tilde{M}_x(y)$ est donnée par

$$\mathbb{E}(\tilde{M}_x(y)) = \frac{1}{2} h_n^2 \kappa [\mu(x) \ddot{F}_x(y) + 2\mu'(x) \dot{F}_x(y)] + o(h_n^2) \quad (1.4)$$

[RL06] ont démontré que la variance de $\tilde{M}_x(y)$ est

$$\text{Var}(\tilde{M}_x(y)) = \frac{1}{nh_n} \kappa \mu(x) F_x(y) (1 - F_x(y)) + o\left(\frac{1}{nh_n}\right). \quad (1.5)$$

On peut également développer $\tilde{M}_x(y)$:

$$\tilde{M}_x(y) = \tilde{M}_x(y) - \mathbb{E}(\tilde{M}_x(y)) + \mathbb{E}(\tilde{M}_x(y)) = O_p(h_n^2 + (nh_n)^{-\frac{1}{2}}).$$

On en conclut que $\left(\frac{1}{\hat{\mu}(x)} - \frac{1}{\mu(x)} \right) \tilde{M}_x(y) = O_p(h_n^2 + (nh_n)^{-\frac{1}{2}})$ et

$$F_{xh}(y) = F_x(y) + \frac{\tilde{M}_x(y)}{\mu(x)} + O_p(h_n^2 + (nh_n)^{-\frac{1}{2}}).$$

□

Lemme 1.1.2 (Espérance de F_{xh} pour Nadaraya-Watson). À l'aide de la Proposition 1.2 et en utilisant les mêmes conditions, [RL06] obtiennent pour les poids de Nadaraya-Watson que

$$\mathbb{E}(F_{xh}(y)) = F_x(y) + \frac{h_n^2 \kappa}{2} \left(\ddot{F}_x(y) + \frac{2\mu'(x) \dot{F}_x(y)}{\mu(x)} \right) + o(1).$$

Si on pose $B_{1x}(y) = \ddot{F}_x(y) + \frac{2\mu'(x)\dot{F}_x(y)}{\mu(x)}$, alors le biais asymptotique de F_{xh} pour ce type de poids est

$$B_x^{(NW)}(y) = \frac{h_n^2 \kappa}{2} B_{1x}(y). \quad (1.6)$$

Selon cette expression, on peut voir clairement que le biais augmente avec h_n .

Dans le cas des poids locaux linéaires et en posant $B_{2x}(y) = \ddot{F}_x(y)$, [GF96] ont démontré que

$$B_x^{(LL)}(y) = \frac{h_n^2 \kappa}{2} B_{2x}(y). \quad (1.7)$$

Pour ce deuxième type de poids, le biais peut être négatif et on utilise donc le biais carré intégré pour comparer les deux estimateurs (NW et LL). Celui-ci est défini de la manière suivante :

$$BI_x = \int_{\text{supp}(Y)} B_x^2(y) dy$$

où $B_x(y)$ est le biais obtenu avec l'un des deux types de poids.

Exemple 1.1.1. Soit $\{x_i\}_{i=1}^{250}$, un échantillon généré à partir de la variable aléatoire $X \sim U(-1, 1)$. Soit $\{e_i\}_{i=1}^{250}$, un échantillon indépendant de $\{x_i\}_{i=1}^{250}$ provenant de la variable aléatoire $E \sim N(0, 1)$. On génère $\{y_i\}_{i=1}^{250}$ à partir du modèle de régression donné par $Y = \sin(X) + E$. Alors,

$$\begin{aligned} F_x(y) &= P(Y \leq y | X = x) \\ &= P(\sin(X) + E \leq y | X = x) \\ &= P(E \leq y - \sin(x)) \\ &= \Phi(y - \sin(x)), \end{aligned}$$

où Φ est la fonction de répartition de la normale centrée réduite. On peut en déduire que

$$\begin{aligned} \dot{F}_x(y) &= -\phi(y - \sin(x)) \cos(x), \\ \ddot{F}_x(y) &= \dot{\phi}(y - \sin(x)) \cos^2(x) + \phi(y - \sin(x)) \sin(x), \end{aligned}$$

où ϕ est la densité de la normale centrée réduite.

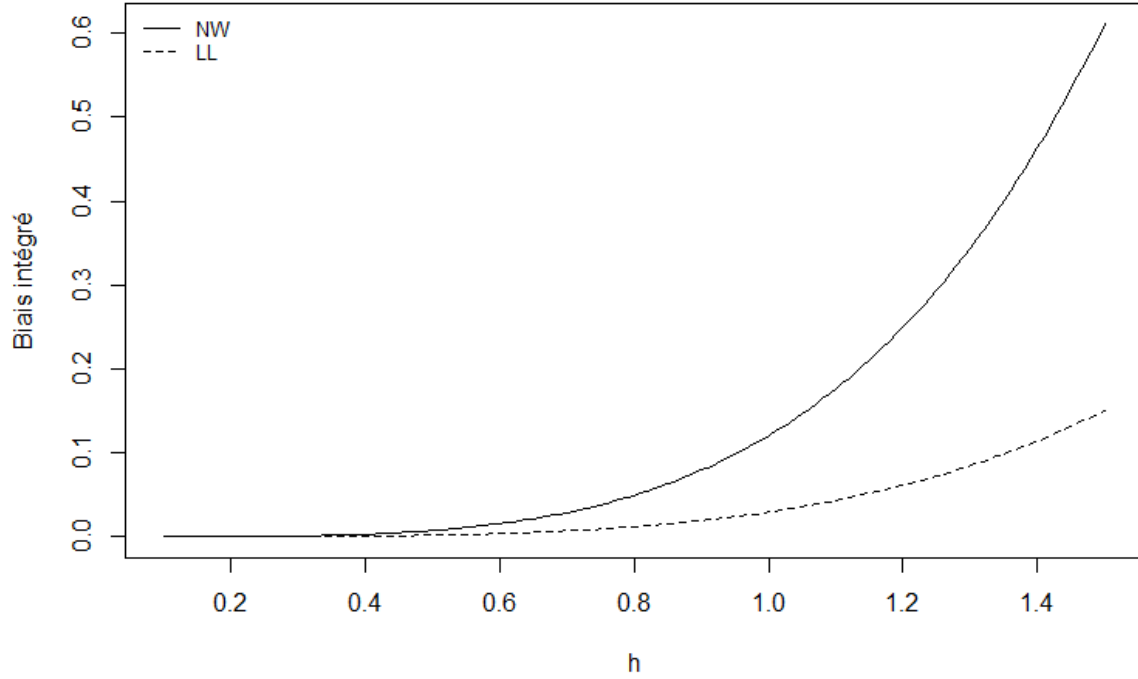


FIGURE 1.3 – Biais intégré de l’estimateur de la distribution conditionnelle de l’Exemple 1.1.1 pour une valeur fixe de x avec les deux types de poids : NW et LL.

La Figure 1.3 illustre le biais intégré de l’estimateur de la distribution conditionnelle en fixant x puis en faisant varier h_n , alors que la Figure 1.4 fixe d’abord h_n et fait varier x .

Dans la Figure 1.3, on constate d’abord que le biais pour LL est toujours plus petit que celui de NW . De plus, le biais des deux estimateurs augmente avec h_n .

On peut voir clairement dans la Figure 1.4 que la différence entre les deux biais augmente lorsqu’on s’éloigne de 0. Le biais pour les poids locaux linéaires est plus petit comparé aux poids de NW lorsqu’on est proche des bornes du support de X .

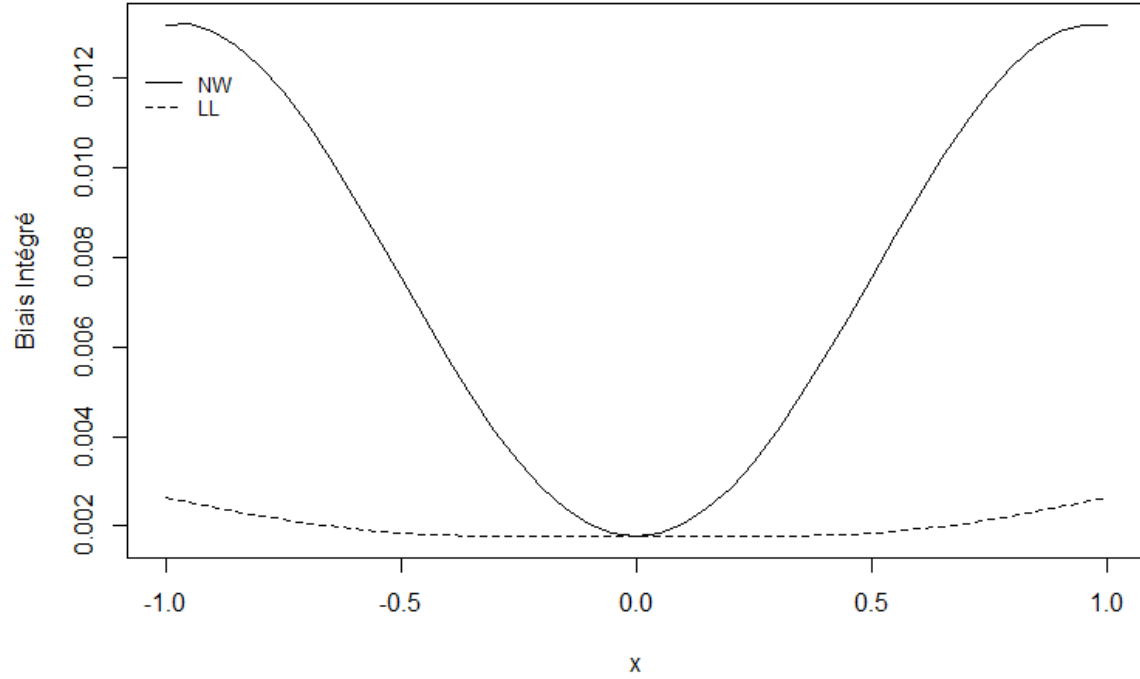


FIGURE 1.4 – Biais intégré de l'estimateur de la distribution conditionnelle de l'Exemple 1.1.1 pour une valeur fixe de h avec les deux types de poids : NW et LL.

La proposition suivante donne la variance de l'estimateur de la distribution conditionnelle F_x .

Proposition 1.3. *Soit (X, Y) un vecteur aléatoire tel que $F_x(y) = P(Y \leq y | X = x)$ satisfait (F1) et (F2). Si h_n est tel que $h_n \rightarrow 0$ et $nh_n \rightarrow \infty$, alors*

$$\text{Var}(F_{xh}(y)) = \frac{1}{nh_n \mu(x)} R(K) \tilde{V}_x(y) + o((nh_n)^{-1}), \quad (1.8)$$

où $\tilde{V}_x(y) = F_x(y)(1 - F_x(y))$.

Démonstration. Voir Théorème 6.1 de [RL06]. □

Cette variance est la même pour les deux types de poids considérés, c.-à.-d. NW et LL. La Figure 1.5 illustre la variance intégrée $VI = \int_{supp(Y)} \text{Var}(F_{xh}(y)) dy$ de l'Exemple 1.1.1. On remarque que la variance diminue lorsque h augmente. Le choix du paramètre de lissage optimal doit faire un compromis entre le biais et la variance de F_{xh} . La section suivante présente la méthode de validation croisée des moindres carrés.

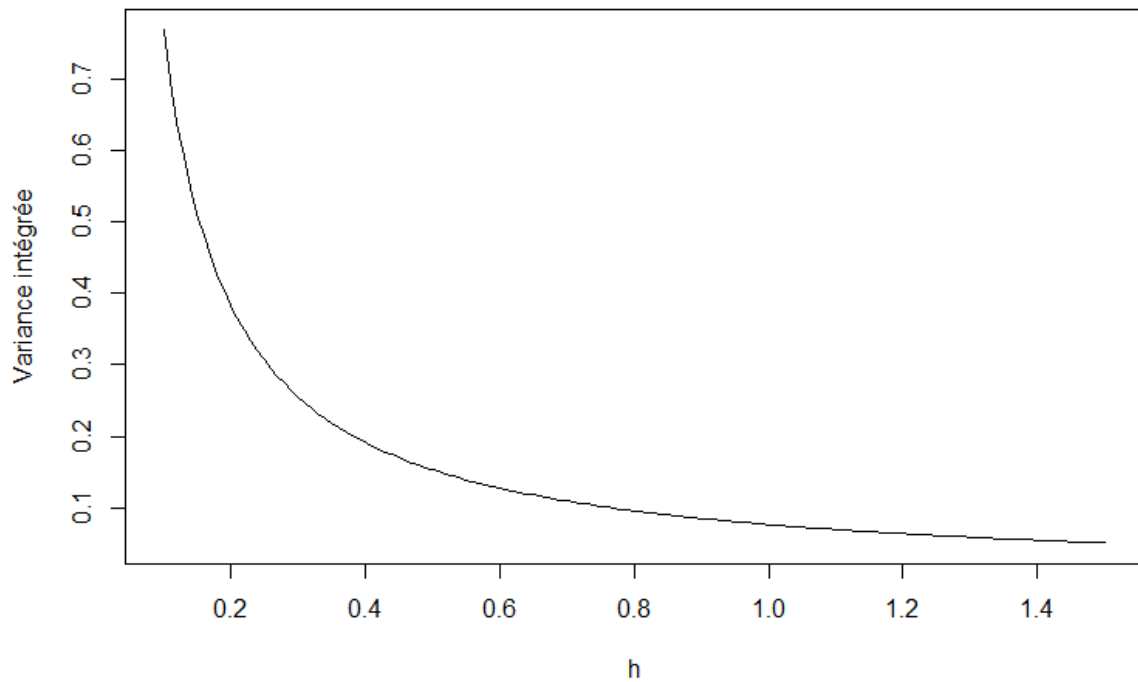


FIGURE 1.5 – Variance intégrée de l'estimateur de la distribution conditionnelle de l'Exemple 1.1.1 pour une valeur fixe de x .

1.1.2 Sélection du paramètre de lissage

Comme il a été mentionné dans le chapitre précédent, le paramètre de lissage joue un rôle important pour la performance de F_{xh} . Pour une petite valeur du paramètre de lissage, le

biais diminue, mais la variance augmente. Pour une grande valeur du paramètre de lissage, la variance diminue, mais le biais augmente. En estimation non-paramétrique, beaucoup de méthodes de sélection du paramètre de lissage se basent sur l'erreur quadratique moyenne (MSE) donnée par

$$\begin{aligned}
MSE(F_{xh}(y)) &= E[(F_{xh}(y) - F_x(y))^2] \\
&= \text{Var}[F_{xh}(y)] + \text{Biais}^2[F_{xh}(y)] \\
&= \frac{1}{nh_n\mu(x)} R(K) \tilde{V}_x(y) + \frac{h_n^4 \kappa^2}{4} B_{lx}^2(y) + o\left(\frac{1}{nh_n} + h_n^4\right), l = 1, 2.
\end{aligned}$$

L'erreur quadratique moyenne asymptotique, notée $AMSE(F_{xh}(y))$, est le terme dominant du $MSE(F_{xh}(y))$.

Cependant, cette erreur est calculée pour une valeur fixe de y . On veut avoir un paramètre h pour l'ensemble des y . Une autre approche consiste à calculer l'erreur quadratique intégrée :

$$ISE(F_{xh}(y)) = \int_{\text{supp}(Y)} [F_{xh}(y) - F_x(y)]^2 dy.$$

Une troisième approche consiste à combiner les deux méthodes en ne gardant que les termes dominants :

$$AMISE(F_{xh}) = \int_{\text{supp}(Y)} AMSE(F_{xh}(y)) dy.$$

La Figure 1.6 illustre l'AMISE de l'estimateur de la distribution conditionnelle avec les deux types de poids de l'Exemple 1.1.1.

On peut déterminer le paramètre de lissage qui minimise l'AMISE :

$$h_{AMISE} = \left(\frac{R(K) \int \tilde{V}_x(y) dy}{\mu(x) \kappa^2 \int \tilde{B}_{lx}^2(y) dy} \right)^{\frac{1}{5}} n^{-\frac{1}{5}},$$

où \tilde{B}_{lx} et \tilde{V}_x sont définies dans les équations (1.6), (1.7) et (1.8) pour $l = 1, 2$.

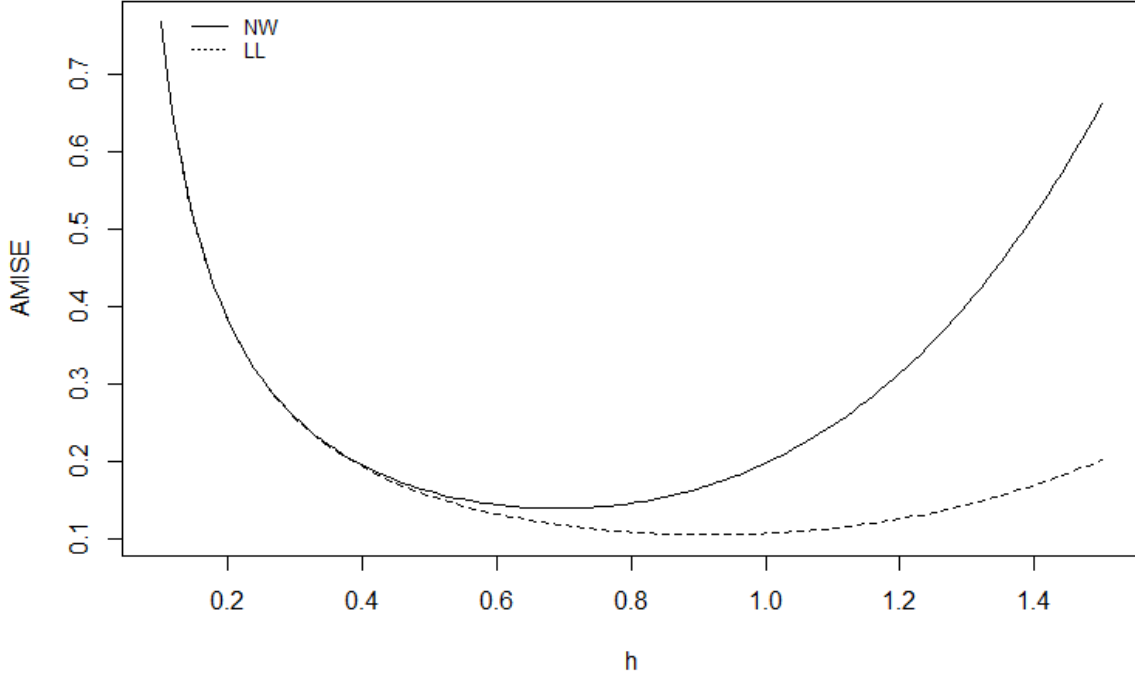


FIGURE 1.6 – AMISE de l’estimateur de la distribution conditionnelle de l’Exemple 1.1.1 pour une valeur fixe de x .

Malheureusement, cette quantité n’est pas connue en pratique puisqu’elle dépend de \tilde{B}_{lx} et \tilde{V}_x qui sont inconnues. Une simple méthode appelée règle du pouce consiste à calculer les quantités inconnues en supposant la normalité du couple (X, Y) (voir [Sil86]). Si la loi de (X, Y) n’est pas proche de la loi normale, cette règle donne des mauvais résultats. La méthode de la validation croisée proposée par [LLR13] est une alternative intéressante qui consiste à minimiser la fonction de perte suivante :

$$CV(h) = \frac{1}{n} \sum_{i=1}^n \int_{\text{supp}(Y)} \{ \mathbf{1}(Y_i \leq y) - F_{X_ih}^{-i}(y) \}^2 w_{ni}(x, h_n) dy,$$

où $F_{X_ih}^{-i}$ est l’estimateur de la fonction de répartition conditionnelle construit en omettant la $i^{\text{ème}}$ observation.

[LLR13] propose de remplacer l'intégrale par une moyenne empirique :

$$CV(h) = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \{ \mathbb{1}(Y_i \leq Y_j) - F_{X_i h}^{-i}(Y_j) \}^2 w_{ni}(x, h_n). \quad (1.9)$$

On considère maintenant deux exemples pour illustrer la minimisation de la fonction de validation croisée pour les deux types de poids. On reprend maintenant l'Exemple 1.1.1 et on utilise la fonction CV proposée par [LLR13].

Suite de l'Exemple 1.1.1

On évalue maintenant la fonction de validation croisée pour un h entre 0.1 et 1.5. La Figure 1.7 utilise les noyaux d'Epanechnikov et triweight ainsi que les deux types de poids pour $x = 0.5$.

Comme l'indique la Figure 1.7, on remarque premièrement l'existence de plusieurs minimums locaux dans le cas du noyau d'Epanechnikov. En pratique, on prend le paramètre h correspondant au plus grand minimum local. Pour un même type de poids, le choix du noyau influence beaucoup le lissage de la fonction de validation croisée. De plus, le noyau triweight semble mieux se comporter et est plus facile à traiter. Selon cette figure, le choix du poids local linéaire semble fournir un paramètre optimal plus grand. On note par h_b et h_d , les deux plus grands minimums locaux des graphiques (b) et (d). Dans cet exemple, on a que $h_b = 0.66$ et $h_d = 1.44$.

On utilise maintenant les paramètres de lissage h_b et h_d pour tracer la fonction de répartition conditionnelle pour les deux types de poids (voir la Figure 1.8).

La courbe en rouge(NW) et celle en bleue(LL) sont très proches de la vraie courbe F_x tracée en noir. Le choix du type de poids ne semble pas trop influencer les résultats. Prenons $x = 0.9$, un point proche des bornes du support de X , et recalculons les paramètres de lissage optimaux, c.-à.-d. qui minimisent la fonction de validation croisée. Comme le représente la Figure 1.9, les deux estimateurs sont plus différents de la vraie fonction de

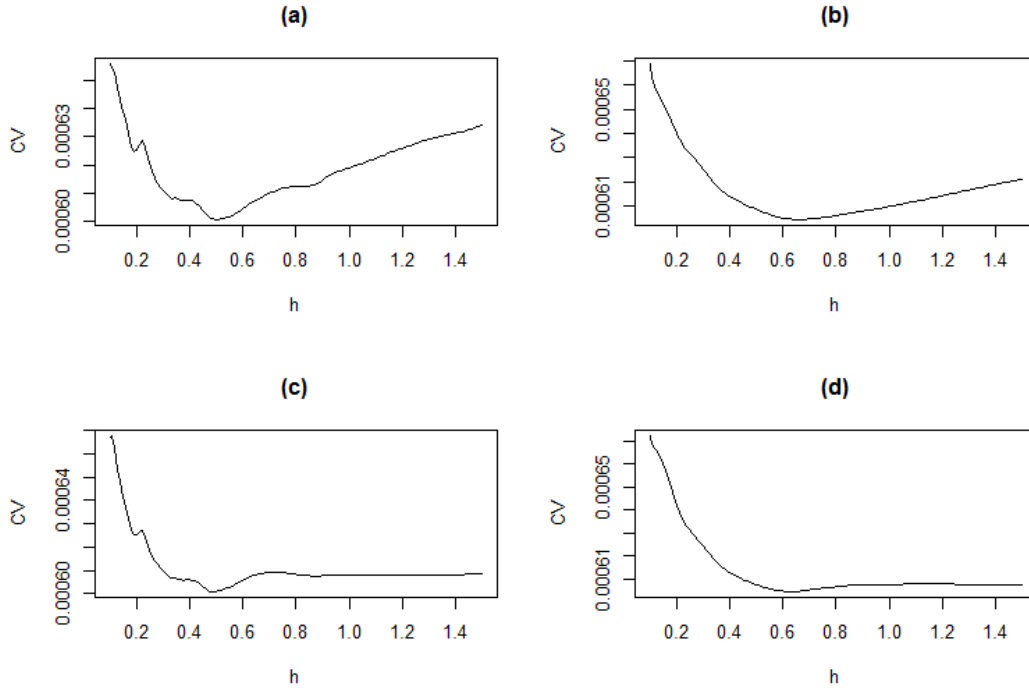


FIGURE 1.7 – Fonctions de validation croisée de l’Exemple 1.1.1 pour $x = 0.5$. Les graphiques (a) et (b) utilisent les poids de Nadaraya-Watson et les noyaux d’Epanechnikov et triweight respectivement, alors que (c) et (d) utilisent les poids locaux linéaires avec les noyaux d’Epanechnikov et triweight respectivement.

répartition conditionnelle que pour $x = 0.5$. Dans cet exemple, l’effet de la covariable X est peu présent. Le prochain exemple illustre un cas extrême.

Exemple 1.1.2. Soit $\{x_i\}_{i=1}^{250}$, un échantillon généré à partir de la variable $X \sim U(-2, 2)$. Soit $\{e_i\}_{i=1}^{250}$, un échantillon provenant de la variable $E \sim N(0, 1)$. On génère $\{y_i\}_{i=1}^{250}$ à partir du modèle de régression donné par $Y = X^2 + E$.

On regarde pour $x = 1.7$, les fonctions de validation croisée (voir Figure 1.10). On remarque que le paramètre de lissage choisi semble être assez petit dans le cas de Nadaraya-Watson. Cette fois ci, $h_b = 0.32$ et $h_d = 0.71$. On trace les deux estimateurs en utilisant h_b et h_d et on compare avec la vraie fonction de répartition. La Figure 1.11 comprend les

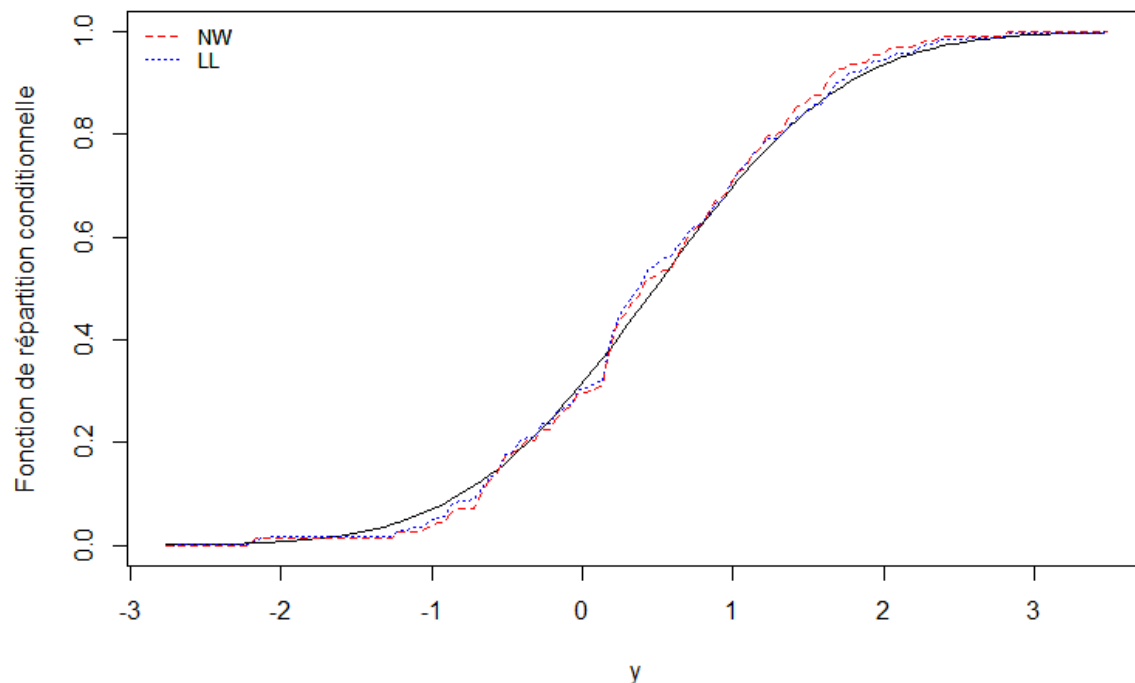


FIGURE 1.8 – Estimation de la distribution conditionnelle de l’Exemple 1.1.1 avec le noyau triweight pour $x = 0.5$ basée sur un échantillon i.i.d. de taille $n = 250$ utilisant les paramètres de lissage obtenues de la minimisation de la fonction de validation croisée.

trois courbes.

L’effet de la covariable dans le modèle de régression joue un rôle très important dans la précision de F_{xh} . Les deux estimateurs ne donnent pas de bons résultats et on pourrait envisager d’utiliser une autre méthode pour estimer la distribution conditionnelle. De nouvelles techniques d’estimation dites pré-ajustées de la fonction de répartition transforment d’abord les données en réduisant l’effet de la covariable. Ces méthodes sont discutées dans l’article de [VGO14]. Comme la précision de F_{xh} n’influence pas beaucoup la force du test d’indépendance que l’on construit au Chapitre 3, on ne considère que ces deux estimateurs.

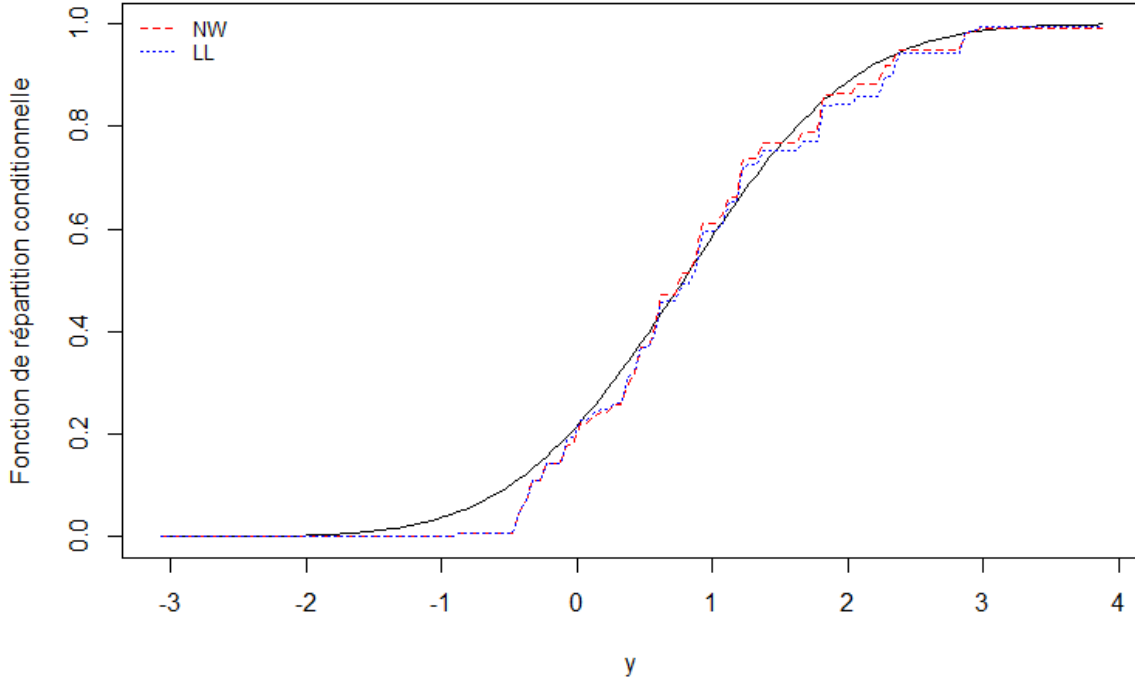


FIGURE 1.9 – Estimation de F_x de l'Exemple 1.1.1 avec le noyau triweight pour $x = 0.9$ basée sur un échantillon i.i.d. de taille $n = 250$ utilisant les paramètres de lissage obtenues de la minimisation de la fonction de validation croisée.

1.1.3 Normalité asymptotique et convergence uniforme de F_{xh} .

Voici un résumé de certaines propriétés de l'estimateur de la distribution conditionnelle.

À l'aide de l'inégalité de Tchebychev (voir [CB01]), on peut démontrer que

$$F_{xh}(y) \xrightarrow{P} F_x(y) \quad \text{pour } y \in \mathbb{R}.$$

Avec le théorème central limite (voir [CB01]), on peut aussi montrer que

$$\sqrt{nh_n}(F_{xh}(y) - F_x(y) - B_x(y)) \xrightarrow{d} N\left(0, \frac{R(K)(F_x(y)(1 - F_x(y)))}{\mu(x)}\right), \quad (1.10)$$

où B_x est le biais pour l'un des deux types de poids.

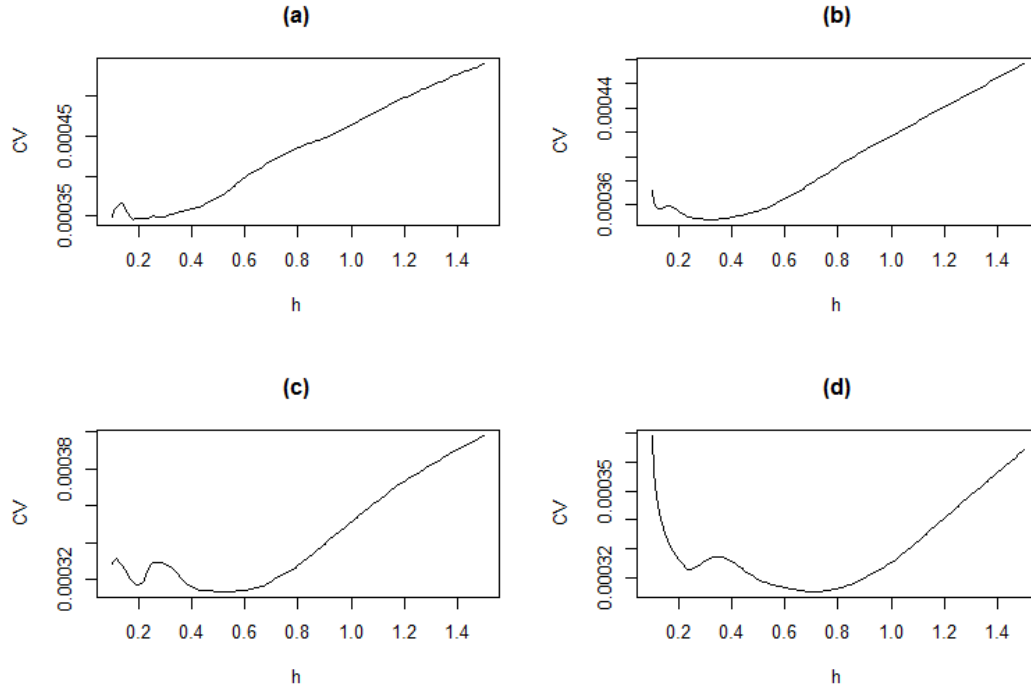


FIGURE 1.10 – Fonctions de validation croisée de l’Exemple 1.1.2 pour $x = 1.7$. Les graphiques (a) et (b) utilisent les poids de Nadaraya-Watson et les noyaux d’Epanechnikov et triweight respectivement, alors que (c) et (d) utilisent les poids locaux linéaires avec les noyaux d’Epanechnikov et triweight respectivement.

De plus, [VOG11] ont démontré que F_{xh} converge uniformément en probabilité vers F_x :

$$\sup_{y \in \mathbb{R}} |F_{xh}(y) - F_x(y)| \xrightarrow{P} 0 \quad \text{quand } n \rightarrow \infty.$$

1.2 Estimation de la copule

1.2.1 Cas non-conditionnel

Soit (Y_1, Y_2) , un couple de variables aléatoires de fonction de répartition H avec marges continues F_1 et F_2 . L’estimateur empirique de la fonction de répartition bidimensionnelle

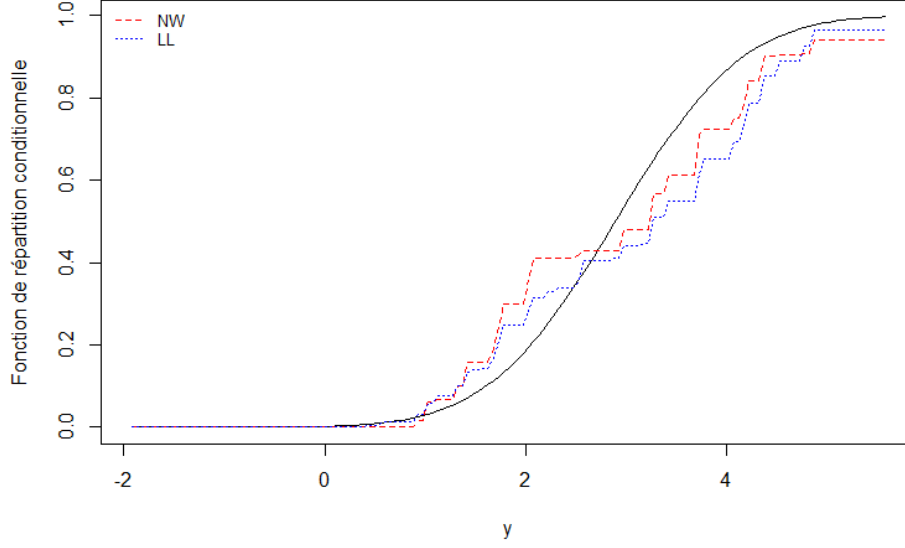


FIGURE 1.11 – Estimation de la distribution conditionnelle de l'Exemple 1.1.2 avec le noyau triweight pour $x = 1.7$. basée sur un échantillon i.i.d. de taille $n = 250$ utilisant les paramètres de lissage obtenues de la minimisation de la fonction de validation croisée.

H est donné par

$$H_n(y_1, y_2) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(Y_{1i} \leq y_1, Y_{2i} \leq y_2) \text{ où } (y_1, y_2) \in \mathbb{R}^2. \quad (1.11)$$

Selon le théorème de Sklar, il existe une copule C unique caractérisant toute la structure de dépendance de (Y_1, Y_2) telle que

$$H(y_1, y_2) = C(F_1(y_1), F_2(y_2)) \quad \forall (y_1, y_2) \in \mathbb{R}.$$

De manière équivalente, on a que

$$C(u, v) = H(F_1^{-1}(u), F_2^{-1}(v)), \quad \forall (u, v) \in [0, 1]^2,$$

où F_1^{-1} et F_2^{-1} sont les fonctions de quantiles. On peut ainsi modéliser la structure de dépendance de façon indépendante de celle des marges. On retrouve une bonne introduc-

tion aux copules dans [Joe14] et [Nel07]. Lorsque Y_1 et Y_2 sont indépendantes, la copule est simplement la suivante : $C_0(u, v) = uv$ appelée copule produit.

On peut estimer la copule C avec l'estimateur non-paramétrique C_n défini comme suit :

$$C_n(u, v) = H_n\{F_{1n}^{-1}(u), F_{2n}^{-1}(v)\} \quad (1.12)$$

$$= \sum_{i=1}^n \mathbf{1}(Y_{1i} \leq F_{1n}^{-1}(u), Y_{2i} \leq F_{2n}^{-1}(v)), \quad (1.13)$$

où F_{1n} et F_{2n} sont les fonctions de répartition empiriques et F_{1n}^{-1} (resp. F_{2n}^{-1}) est la fonction inverse généralisée de F_{1n} (resp. F_{2n}) définie pour $u \in [0, 1]$ par

$$F_{jn}^{-1}(u) = \inf\{y \in \mathbb{R} \mid F_{jn}(y) \geq u\} \quad j = 1, 2. \quad (1.14)$$

C_n est appelée la copule empirique. Cet estimateur a la particularité d'être asymptotiquement sans biais.

Lorsque $C(u, v) = uv$, c-à-d la copule d'indépendance, le comportement asymptotique du processus $\mathbb{C}_n^0 = \sqrt{n}(C_n - C_0)$ a été étudié par [Deh79]. En fixant $(u, v) \in (0, 1)$, on a la convergence en loi de la suite de variable aléatoire suivante :

$$\sqrt{n}(C_n(u, v) - C_0(u, v)) \xrightarrow{d} N(0, uv(u-1)(v-1)).$$

Un simple test d'indépendance peut prendre la forme suivante :

$$H_0 : C = C_0 \quad \text{contre} \quad H_1 : C \neq C_0.$$

Contrairement à des tests d'égalité de moyennes, on cherche ici à tester l'égalité entre deux fonctions. L'étude du comportement asymptotique de \mathbb{C}_n^0 comme une suite de fonctions aléatoires de $l^\infty([0, 1]^2)$ peut nous aider à construire le test d'hypothèse. Pour mieux présenter ce concept, on définit le pont brownien.

Définition 1.2.1 (Pont brownien). Un pont brownien bidimensionnel, noté \mathbb{B} , est une fonction aléatoire définie sur $[0, 1]^2$ à trajectoires continues telle que pour $(u, v) \in (0, 1)^2$, on a

- $\mathbb{B}(u, v) \sim N(0, uv(1 - uv))$,
- $\text{Cov}(\mathbb{B}(u, v), \mathbb{B}(u', v')) = \min(u', u')\min(v, v') - uu'vv'$.

[Deh79] et [Stu86] ont montré que la suite \mathbb{C}_n^0 converge faiblement dans $l^\infty([0, 1]^2)$ vers un processus \mathbb{G}_0 :

$$\mathbb{C}_n^0 = \sqrt{n}(C_n - C_0) \rightsquigarrow \mathbb{G}_0, \quad (1.15)$$

où $\mathbb{G}_0(u, v) = \mathbb{B}(u, v) - v\mathbb{B}(u, 1) - u\mathbb{B}(1, v)$ avec \mathbb{B} , un pont brownien.

Ce résultat a été utilisé dans les articles de [Deh80] et [Deh81] pour la construction d'un test d'indépendance basé sur \mathbb{C}_n^0 . [GR04] ont étendu ces résultats au cas sériel.

1.2.2 Cas conditionnel

Dans un contexte conditionnel, soit (Y_1, Y_2, X) , un triplet aléatoire prenant des valeurs dans \mathbb{R}^3 . Dénotons par $H_x(y_1, y_2) = P(Y_1 \leq y_1, Y_2 \leq y_2 | X = x)$, la fonction de répartition conjointe de (Y_1, Y_2) sachant $X = x$. De plus, les fonctions de répartition de $Y_1 | X = x$ et de $Y_2 | X = x$ sont données par

$$F_{1x}(y) = P(Y_1 \leq y | X = x), \quad F_{2x}(y) = P(Y_2 \leq y | X = x).$$

Lorsque F_{1x} et F_{2x} sont continues, le théorème de Sklar (revue par [Pat06]) nous garantit l'existence d'une copule unique C_x telle que

$$H_x(y_1, y_2) = C_x\{F_{1x}^{-1}(u), F_{2x}^{-1}(v)\} \quad \forall (y_1, y_2) \in \mathbb{R}^2, \quad (1.16)$$

ou de façon équivalente :

$$C_x(u, v) = H_x\{F_{1x}^{-1}(u), F_{2x}^{-1}(v)\} \quad \forall (u, v) \in [0, 1]^2. \quad (1.17)$$

La copule conditionnelle fournit donc un outil efficace pour modéliser la dépendance entre Y_1 et Y_2 sachant $X = x$. L'exemple suivant permet de bien visualiser comment la dépendance entre Y_1 et Y_2 peut être influencée par les valeurs de X .

Exemple 1.2.1. Supposons qu'on génère un échantillon $\{x_i\}_{i=1}^{500}$ d'une loi uniforme $U(-1, 1)$. On prend ensuite un échantillon $\{u_{1i}, u_{2i}\}_{i=1}^{500}$ en utilisant une copule gaussienne où le paramètre de dépendance θ_i est donné par $\theta_i = x_i$. On génère ensuite $\{y_{1i}\}_{i=1}^{500}$ et $\{y_{2i}\}_{i=1}^{500}$ en prenant le quantile d'une loi normale centrée réduite des u_{1i} et u_{2i} . La Figure 1.10 illustre les couples d'observations $\{(y_{1i}, y_{2i})\}_{i=1}^{500}$. On décide de regrouper ensemble les observations $\{(y_{1i}, y_{2i}, x_i)\}$ telles que $|x_i| \leq 0.2$. Le graphique A de la Figure 1.11 illustre ces observations. Ce graphique montre un nuage de points presque uniforme et indépendant. On sélectionne ensuite les observations telles que $|x_i| \geq 0.8$ et on les illustrent dans le graphique B. Dans celui-ci, on voit une forte corrélation entre Y_1 et Y_2 .

On peut voir que la dépendance n'est pas la même dans les deux groupes. La valeur de la covariable influence donc le lien entre Y_1 et Y_2 .

Soit $\{Y_{1i}, Y_{2i}, X_i\}_{i=1}^n$, un échantillon i.i.d. Pour utiliser les estimateurs des fonctions de répartition conditionnelle de $Y_1|X = x$ et $Y_2|X = x$, on prend les notations suivantes :

$$F_{1xh}(y) = \sum_{i=1}^n \mathbb{1}(Y_{1i} \leq y) w_{ni}(x, h_n), \quad F_{2xh}(y) = \sum_{i=1}^n \mathbb{1}(Y_{2i} \leq y) w_{ni}(x, h_n).$$

Pour en venir à un estimateur de la copule conditionnelle, on utilise l'estimateur de H_x donné par

$$H_{xh}(y_1, y_2) = \sum_{i=1}^n \mathbb{1}(Y_{1i} \leq y_1, Y_{2i} \leq y_2) w_{ni}(x, h_n). \quad (1.18)$$

[VOG11] ont proposé et étudié l'estimateur de la copule conditionnelle basé sur H_{xh} donné par

$$\begin{aligned} C_{xh}(u_1, u_2) &= H_{xh} \{F_{1xh}^{-1}(u_1), F_{2xh}^{-1}(u_2)\} \\ &= \sum_{i=1}^n \mathbb{1}(Y_{1i} \leq F_{1xh}^{-1}(u_1), Y_{2i} \leq F_{2xh}^{-1}(u_2)) w_{ni}(x, h_n), \end{aligned} \quad (1.19)$$

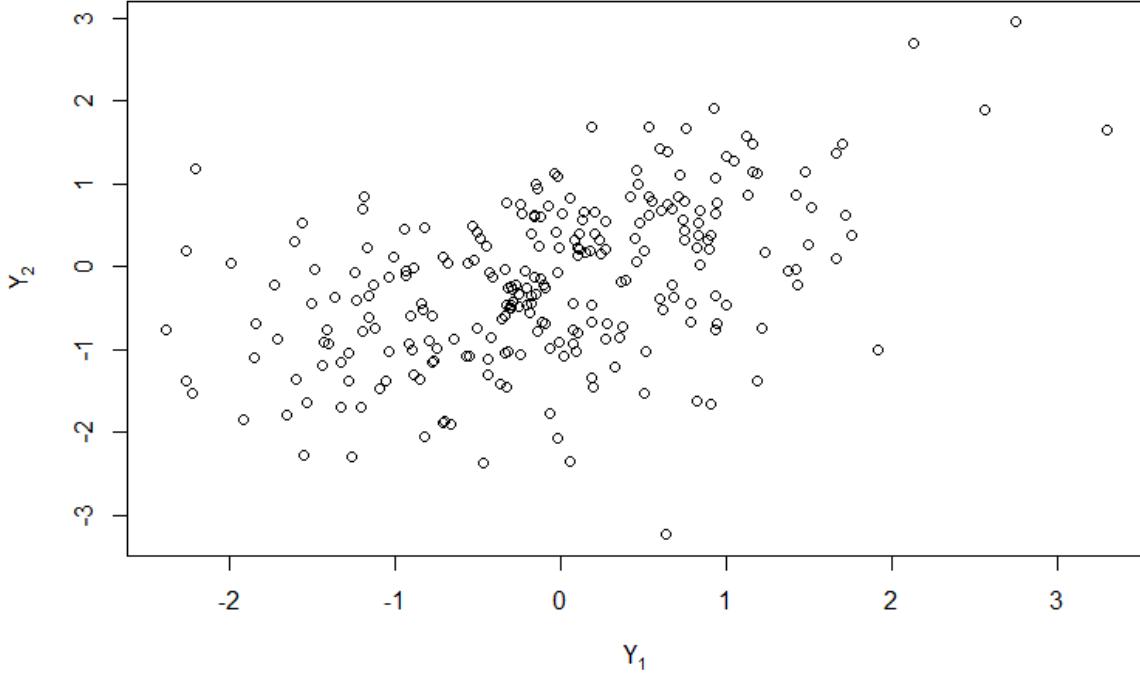


FIGURE 1.12 – Graphique des observations $\{(y_{1i}, y_{2i})\}_{i=1}^n$ de l'Exemple 1.2.1.

où $(u_1, u_2) \in [0, 1]^2$ et $F_{kxh}^{-1}(u) = \inf\{y \in \mathbb{R} \mid F_{kxh}(y) \geq u\}$ est l'inverse généralisé de F_{kxh} pour $k = 1, 2$.

Maintenant, pour discuter des propriétés de C_{xh} , on a besoin de certaines notations et hypothèses sur la copule. Soit $\dot{C}_x(u, v) = \frac{\partial}{\partial x} C(u, v)$, $C_x^{(1)}(u, v) = \frac{\partial}{\partial u} C_x(u, v)$, $C_x^{(2)}(u, v) = \frac{\partial}{\partial v} C_x(u, v)$, $\dot{C}_x^{(j)}(u_1, u_2) = \frac{\partial}{\partial x \partial u_j} C_x(u_1, u_2)$ et $C_x^{(i,j)}(u_1, u_2) = \frac{\partial^2}{\partial u_i \partial u_j} C(u_1, u_2)$ pour $i, j = 1, 2$.

Hypothèses (C)

(C1) $z \rightarrow \dot{C}_z$ est uniformément continue dans un voisinage de x ,

(C2) $z \rightarrow C_z^{(1)}$ est uniformément continue dans un voisinage de x ,

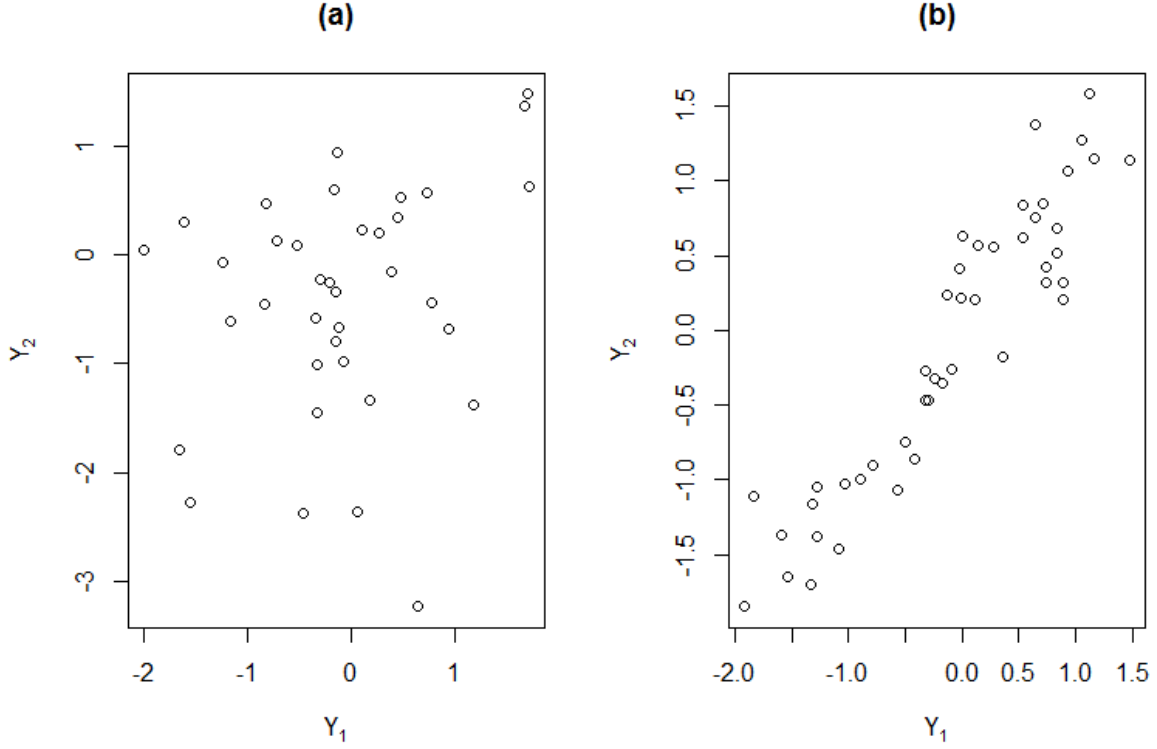


FIGURE 1.13 – Graphique des observations $\{(y_{1i}, y_{2i})\}_{i=1}^n$ séparées en deux groupes de l'Exemple 1.2.1 : (a) $|x_i| \leq 0.2$ et (b) $|x_i| \geq 0.8$.

(C3) $z \rightarrow C_z^{(2)}$ est uniformément continue dans un voisinage de x .

En utilisant ces hypothèses, [GVO10] ont démontré que le biais asymptotique de H_{xh} est donnée par :

$$\begin{aligned}
 B_{H_x}(u, v) &= \ddot{C}_x(u, v) + 2\dot{C}_x^{(1)}(u, v)\dot{F}_{1x}\{F_{1x}^{-1}(u)\} + 2\dot{C}_x^{(2)}(u, v)\dot{F}_{2x}\{F_{2x}^{-1}(v)\} \\
 &\quad C_x^{(1,1)}(u, v)[\dot{F}_{1x}\{F_{1x}^{-1}(u)\}]^2 + C_x^{(2,2)}(u, v)[\dot{F}_{2x}\{F_{2x}^{-1}(v)\}]^2 \\
 &\quad + 2C_x^{(1,2)}(u, v)\dot{F}_{1x}\{F_{1x}^{-1}(u)\}\dot{F}_{2x}\{F_{2x}^{-1}(v)\}.
 \end{aligned}$$

Lorsque qu'on a l'indépendance conditionnelle de $Y_1|X = x$ et $Y_2|X = x$ (c-à-d $C_x(u, v) = uv$), ce biais asymptotique possède beaucoup de dérivées qui valent zéro. La seule dérivée non-nulle est $C_x^{(1,2)}$ qui vaut 1. On peut donc dire que le biais asymptotique sous H_0 est

donné par

$$\tilde{B}_{H_x}(u, v) = 2\dot{F}_{1x}\{F_{1x}^{-1}(u)\}\dot{F}_{2x}\{F_{2x}^{-1}(v)\}.$$

Pour bâtir un test d'indépendance conditionnelle basé sur C_x , on utilise le processus suivant :

$$\mathbb{C}_{xh}^0 = \sqrt{nh_n}\{C_{xh} - C_0\}. \quad (1.20)$$

Théorème 1.1. *Du résultat général de [VOG11], on peut retrouver la convergence faible sous l'hypothèse de l'indépendance conditionnelle :*

$$\mathbb{C}_{xh}^0 \rightsquigarrow \mathbb{G}_{xh}^0 = V_{x0}\mathbb{G}_0 + E_{x0}\tilde{B}_{H_x} \quad (1.21)$$

où

$$\begin{aligned} V_{x0} &= \sqrt{\lim_{n \rightarrow \infty} nh_n \sum_{i=1}^n w_{ni}^2(x, h_n)}, \\ E_{x0} &= \lim_{n \rightarrow \infty} \sqrt{nh_n} \sum_{i=1}^n w_{ni}(x, h_n)(x_i - x)^2, \\ \tilde{B}_{H_x} &= 2\dot{F}_{1x}\{F_{1x}^{-1}(u)\}\dot{F}_{2x}\{F_{2x}^{-1}(v)\} \quad (\text{Biais asymptotique de } H_{xh} \text{ sous } H_0) \end{aligned}$$

Nous allons voir dans le chapitre quatre que pour faire un test d'indépendance conditionnelle, l'une des techniques consiste à obtenir plusieurs copies du processus en utilisant la représentation asymptotique (1.21). Pour obtenir des répliques valides, on doit d'abord être en mesure d'estimer les quantités inconnues de l'équation (1.21). Pour V_{x0} et E_{x0} , on peut utiliser $V_{xh} = \sqrt{\sum_{i=1}^n nh_n w_{ni}^2(x, h_n)}$ et $E_{xh} = \sqrt{nh_n} \sum_{i=1}^n w_{ni}(x, h_n)(x_i - x)^2$. On propose $\tilde{B}_{H_{xh}}(u, v) = 2\dot{F}_{1xh}\{F_{1xh}^{-1}(u)\}\dot{F}_{2xh}\{F_{2xh}^{-1}(v)\}$ comme estimateur de \tilde{B}_{H_x} où $\dot{F}_{1xh}(y) = \sum_{i=1}^n \mathbb{1}(Y_{1i} \leq y) \dot{w}_{ni}(x)$. Au chapitre suivant, nous allons démontrer certaines propriétés de \dot{F}_{xh} et prouver que $\tilde{B}_{H_{xh}} \xrightarrow{P} \tilde{B}_{H_x}$.

Avant de passer au prochain chapitre, on se penche sur des méthodes de sélection du paramètre de lissage de C_{xh} . On développe premièrement la validation croisée de [LLR13]

dans le cas des copules conditionnelles. Ensuite, on présente une méthode seulement valable sous H_0 (c-à-d quand $C_x(u, v) = uv$).

1.2.3 Sélection du paramètre de lissage

Validation croisée de [LLR13]

La méthode utilisée dans ce chapitre pour trouver le paramètre de lissage pour une fonction de répartition unidimensionnelle est celle de [LLR13]. Leur article généralise cette technique dans le cas multidimensionnel. On peut donc l'adapter aux copules conditionnelles et chercher à minimiser la fonction suivante :

$$CV^*(h) = \sum_{i=1}^n \int_0^1 \int_0^1 \left\{ \mathbb{1}(Y_i \leq F_{1xh}^{-1}(u), Y_2 \leq F_{2xh}^{-1}(v)) - C_{X_ih}^{(-i)}(u, v) \right\}^2 w_{ni}(x, h_n) du dv \quad (1.22)$$

En pratique, on sait que calculer l'inverse généralisé demande beaucoup d'opérations. On commence par calculer les pseudos-observations centrées en x :

$$(U_{1i}, U_{2i}) = (F_{1xh}(Y_{1i}), F_{2xh}(Y_{2i})) \quad \forall i = 1 \dots n.$$

$$CV^*(h) = \sum_{i=1}^n \int_0^1 \int_0^1 \left\{ \mathbb{1}(U_{1i} \leq u, U_{2i} \leq v) - C_{X_ih}^{(-i)}(u, v) \right\}^2 w_{ni}(x, h_n) du dv.$$

On peut encore aller plus loin et développer $C_{X_ih}^{-i}$ et refaire la même chose :

$$(\tilde{U}_{1j}^{(i)}, \tilde{U}_{2j}^{(i)}) = (F_{1X_ih}(Y_{1j}), F_{2X_ih}(Y_{2j})) \quad \forall j = 1 \dots n.$$

$$\begin{aligned} CV^*(h) &= \sum_{i=1}^n \int_0^1 \int_0^1 \left\{ \mathbb{1}(U_{1i} \leq u, U_{2i} \leq v) \right. \\ &\quad \left. - \sum_{j \neq i} \mathbb{1}(\tilde{U}_{1j}^{(i)} \leq u, \tilde{U}_{2j}^{(i)} \leq v) w_{nj}^{(-i)}(x_i) \right\}^2 w_{ni}(x, h_n) du dv \\ &= \sum_{i=1}^n w_{ni}(x, h_n) \left[I_1^{(i)} + \sum_{j \neq i} w_{nj}^{-i}(x_i) \left\{ [-2I_2^{(i,j)} + \sum_{l \neq i} w_{nl}^{(-i)}(x_i) I_3^{(i,j,l)}] \right\} \right], \end{aligned}$$

où

1. $I_1^{(i)} = \int_0^1 \int_0^1 \mathbb{1}(U_{1i} \leq u, U_{2i} \leq v) du dv,$
2. $I_2^{(i,j)} = \int_0^1 \int_0^1 \mathbb{1}(U_{1i} \leq u, U_{2i} \leq v) \mathbb{1}(\tilde{U}_{1j}^{(i)} \leq u, \tilde{U}_{2j}^{(i)} \leq v) du dv,$
3. $I_3^{(i,j,l)} = \int_0^1 \int_0^1 \mathbb{1}(\tilde{U}_{1j}^{(i)} \leq u, \tilde{U}_{2j}^{(i)} \leq v) \mathbb{1}(\tilde{U}_{1l}^{(i)} \leq u, \tilde{U}_{2l}^{(i)} \leq v) du dv.$

L'un des avantages de cette méthode est le fait qu'on peut calculer facilement les intégrales précédentes :

$$\begin{aligned}
I_1^{(i)} &= \int_0^1 \mathbb{1}(U_{2i} \leq v) \int_0^1 \mathbb{1}(U_{1i} \leq u) du dv \\
&= \int_{U_{2i}}^1 \int_{U_{1i}}^1 du dv \\
&= (1 - U_{1i})(1 - U_{2i}).
\end{aligned}$$

$$\begin{aligned}
I_2^{(i,j)} &= \int_0^1 \mathbb{1}(U_{1i} \leq u, \tilde{U}_{1j}^{(i)} \leq u) \int_0^1 \mathbb{1}(U_{2i} \leq v, \tilde{U}_{2j}^{(i)} \leq v) du dv \\
&= \int_0^1 \mathbb{1}(\max\{U_{1i}, \tilde{U}_{1j}^{(i)}\} \leq u) \int_0^1 \mathbb{1}(\max\{U_{2i}, \tilde{U}_{2j}^{(i)}\} \leq v) du dv \\
&= \left(1 - \max\{U_{1i}, \tilde{U}_{1j}^{(i)}\}\right) \left(1 - \max\{U_{2i}, \tilde{U}_{2j}^{(i)}\}\right)
\end{aligned}$$

$$\begin{aligned}
I_3^{(i,j,l)} &= \int_0^1 \mathbb{1}(\tilde{U}_{1j}^{(i)} \leq u, \tilde{U}_{1l}^{(i)} \leq u) \int_0^1 \mathbb{1}(\tilde{U}_{2j}^{(i)} \leq v, \tilde{U}_{2l}^{(i)} \leq v) du dv \\
&= \int_0^1 \mathbb{1}(\max\{\tilde{U}_{1j}^{(i)}, \tilde{U}_{1l}^{(i)}\} \leq u) \int_0^1 \mathbb{1}(\max\{\tilde{U}_{2j}^{(i)}, \tilde{U}_{2l}^{(i)}\} \leq v) du dv \\
&= \left(1 - \max\{\tilde{U}_{1j}^{(i)}, \tilde{U}_{1l}^{(i)}\}\right) \left(1 - \max\{\tilde{U}_{2j}^{(i)}, \tilde{U}_{2l}^{(i)}\}\right).
\end{aligned}$$

Un critère que l'on pourrait utiliser est le suivant :

$$\begin{aligned}
CV^*(h) = & \sum_{i=1}^n w_{ni}(x, h_n) \left[(1 - U_{1i})(1 - U_{2i}) \right. \\
& + \sum_{j \neq i} w_{nj}^{(-i)}(x_i) \{ -2(1 - \max\{U_{1i}, \tilde{U}_{1j}^{(i)}\})(1 - \max\{U_{2i}, \tilde{U}_{2j}^{(i)}\}) \\
& \left. + \sum_{l \neq i} w_{nl}^{(-i)}(x_i) (1 - \max\{\tilde{U}_{1j}^{(i)}, \tilde{U}_{1l}^{(i)}\})(1 - \max\{\tilde{U}_{2j}^{(i)}, \tilde{U}_{2l}^{(i)}\}) \right]. \quad (1.23)
\end{aligned}$$

Il resterait à tester si ce critère fonctionne bien à travers des simulations. La technique suivante est basée sur la statistique de Cramér-von Mises.

La méthode Cramér-von Mises

Elle consiste à calculer le paramètre h minimisant la distance de Cramér-von Mises donnée par

$$\int_0^1 \int_0^1 (C_{xh}(u, v) - uv)^2 du dv.$$

On peut développer ce terme et enlever le terme qui ne dépend pas de h , on trouve

$$\int_0^1 \int_0^1 C_{xh}^2(u, v) - 2uvC_{xh}(u, v) du dv.$$

On utilise la même transformation que dans la méthode précédente :

$$(U_{1i}, U_{2i}) = (F_{1xh}(Y_{1i}), F_{2xh}(Y_{2i})) \quad \forall i = 1 \dots n.$$

$$\begin{aligned}
P(h) = & \int_0^1 \int_0^1 \left[\sum_{i=1}^n \sum_{j=1}^n \mathbb{1}(U_{1i} \leq u, U_{2i} \leq v) \mathbb{1}(U_{1j} \leq u, U_{2j} \leq v) w_{ni}(x, h_n) w_{nj}(x) \right. \\
& \left. - 2uv \sum_{i=1}^n \mathbb{1}(U_{1i} \leq u, U_{2i} \leq v) w_{ni}(x, h_n) \right] du dv \\
= & \sum_{i=1}^n w_{ni}(x, h_n) \left\{ \sum_{j=1}^n w_{nj}(x) [1 - \max\{U_{1i}, U_{1j}\}] [1 - \max\{U_{2i}, U_{2j}\}] \right. \\
& \left. - \frac{1}{2} [1 - U_{1i}] [1 - U_{2i}] \right\}
\end{aligned}$$

Cette deuxième méthode demande beaucoup moins de temps à minimiser pour des valeurs fixes de h . Elle requiert certaines modifications pour mieux fonctionner. Pour le test d'indépendance, on utilisera un paramètre de lissage simple à calculer donné par $h = \bar{\sigma} n^{-\frac{1}{5}}$ où $\bar{\sigma}$ est l'étendue interquartile de $\{X_i\}_{i=1}^n$.

Dans le prochain chapitre, on développe un estimateur de la dérivée de la distribution conditionnelle pour nous permettre d'estimer le biais asymptotique de H_x sous H_0 .

Chapitre 2

Estimation de la dérivée de la distribution conditionnelle

2.1 Estimateur à noyau de la dérivée de la fonction de distribution conditionnelle

Jusqu'à présent, on travaillait avec un design aléatoire où $(Y, X) \sim H$ et on avait que $Y|X = x \sim F_x$ avec F_x . En pratique, pour faire les calculs, on prenait les poids de Nadaraya-Watson et on devait utiliser l'espérance conditionnelle en présence de (Y, X) (voir la Proposition 1.2). Le problème est qu'on doit toujours spécifier le type de poids pour faire les calculs et qu'on ne veut pas refaire les mêmes développements pour d'autres poids. Dans ce chapitre, on décide donc de considérer une deuxième approche : le design fixe. Voici les principales différences entre les deux designs :

Design aléatoire	Design fixe
1. $\{X_i\}_{i=1}^n$ ont pour densité μ	1. $\{x_i\}_{i=1}^n$ fixes
2. $Y_i X_i = x \sim F_x$	2. $Y_i \sim F_{x_i}$
3. (Y_i, X_i) sont i.i.d.	3. Y_i sont indépendantes.

Pour certaines démonstrations, on fait des développements utilisant Nadaraya-Watson dans un cadre aléatoire, mais la suite du mémoire est fait dans un design fixe. On peut passer d'un design à l'autre avec des techniques similaires au chapitre deux, c.-à.-d. en conditionnant d'abord par la covariable et en faisant un développement en série de Taylor autour de $X = x$. Pour estimer la dérivée par rapport à x de la fonction de répartition conditionnelle F_x , on propose d'utiliser l'estimateur :

$$\begin{aligned}\dot{F}_{xh}(y) &= \frac{\partial}{\partial x} F_{xh}(y) \\ &= \sum_{i=1}^n \mathbb{1}(Y_{1i} \leq y) \dot{w}_{ni}(x, h_n),\end{aligned}$$

où $\dot{w}_{ni}(x, h_n) = \frac{d}{dx} w_{ni}(x, h_n)$. On donne maintenant certaines hypothèses pour démontrer la convergence en loi et la convergence en probabilité :

Hypothèse (H)

$$(H) \quad h_n = O(n^{-\frac{1}{7}}).$$

Hypothèses (F)

(F1) \dot{F}_x est continue en x ,

(F2) $\ddot{F}_x = \frac{\partial}{\partial x} \dot{F}_x$ est continue dans un voisinage de x .

Les conditions suivantes sont énoncées dans un design fixe, mais peuvent être converties dans un design aléatoire en ajoutant un p aux notations "O" et "o".

Hypothèses (P)

$$(P1) \quad \sum_{i=1}^n \dot{w}_{ni}(x, h_n) = o(1),$$

$$(P2) \quad \sum_{i=1}^n \dot{w}_{ni}(x, h_n)(x_i - x) - 1 = o(1),$$

$$(P3) \sum_{i=1}^n \dot{w}_{ni}(x, h_n)(x_i - x)^2 = o(1),$$

$$(P4) \sum_{i=1}^n \dot{w}_{ni}^2(x, h_n) = O\left(\frac{1}{nh_n^3}\right),$$

$$(P5) \sum_{i=1}^n \dot{w}_{ni}^2(x, h_n)(x_i - x) = o\left(\frac{1}{nh_n^3}\right).$$

Remarque. L'hypothèse (H) sur le paramètre de lissage joue un rôle important sur l'ordre du biais et la variance asymptotiques. Les hypothèses (F) contrôlent le comportement des dérivées de F_x autour de x . Les hypothèses (P) concernent le système de poids utilisé.

Avant de vérifier les hypothèses pour les poids de Nadaraya-Watson, on a besoin de certains nouveaux résultats sur les noyaux.

Proposition 2.1. *Soit $\{X_i\}_{i=1}^n$ un échantillon i.i.d. tiré de X ayant pour densité μ deux fois différentiable et K , un noyau d'ordre deux défini sur $[-1, 1]$. Soit h_n tel que $h_n \rightarrow 0$ et $nh_n^3 \rightarrow \infty$. On a que*

1. $\frac{1}{nh_n^2} \sum_{i=1}^n \dot{K}\left(\frac{X_i - x}{h_n}\right) \xrightarrow{P} (2K(1) - 1)\mu'(x),$
2. $\frac{1}{nh_n^2} \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right) \dot{K}\left(\frac{X_i - x}{h_n}\right) (X_i - x) \xrightarrow{P} \frac{1}{2}(2K^2(1) - R(K))\mu'(x),$
3. $\frac{1}{nh_n^2} \sum_{i=1}^n \dot{K}\left(\frac{X_i - x}{h_n}\right) (X_i - x) \xrightarrow{P} (2K(1) - 1)\mu(x),$
4. $\frac{1}{nh_n^4} \sum_{i=1}^n \dot{K}\left(\frac{X_i - x}{h_n}\right) (X_i - x)^2 \xrightarrow{P} (2K(1) - 3\kappa)\mu'(x),$
5. $\frac{1}{nh_n} \sum_{i=1}^n \left\{ \dot{K}\left(\frac{X_i - x}{h_n}\right) \right\}^2 \xrightarrow{P} R(\dot{K})\mu(x),$
6. $\frac{1}{nh_n^3} \sum_{i=1}^n \left\{ \dot{K}\left(\frac{X_i - x}{h_n}\right) \right\}^2 (X_i - x) \xrightarrow{P} D\mu'(x) \text{ où } D \in \mathbb{R}^*.$

Proposition 2.2. *Les hypothèses (P) sont vérifiées pour les poids de Nadaraya-Watson.*

Démonstration. On fait la démonstration pour un design aléatoire. Les poids de Nadaraya-Watson sont donnés par

$$w_{ni}(x, h_n) = \frac{K\left(\frac{X_i - x}{h_n}\right)}{S_n} \text{ où } S_n = \sum_{j=1}^n K\left(\frac{X_j - x}{h_n}\right).$$

Dans le Lemme 1.1.1, on a défini

$$\hat{\mu}(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right)$$

pour un design aléatoire et on a démontré dans la Remarque 1.1 que $\hat{\mu}(x) = \mu(x) + o_p(1)$. En constatant que $S_n = nh_n \hat{\mu}(x)$, on obtient que $S_n = O_p(nh_n)$. Dans les calculs qui suivent, on utilise à profusion les résultats des Propositions 1.1 et 2.1.

On dérive maintenant $w_{ni}(x, h_n)$ par rapport à x et on trouve que

$$\dot{w}_{ni}(x, h_n) = \frac{1}{h_n S_n^2} \left[-S_n \dot{K}\left(\frac{X_i - x}{h_n}\right) + K\left(\frac{X_i - x}{h_n}\right) \sum_{j=1}^n \dot{K}\left(\frac{X_j - x}{h_n}\right) \right]. \quad (2.1)$$

Pour vérifier (P1), on sait que $\sum_{i=1}^n w_{ni}(x, h_n) = 1$. En prenant la dérivée par rapport à x , on obtient que $\sum_{i=1}^n \dot{w}_{ni}(x, h_n) = 0$.

On s'intéresse maintenant à la deuxième hypothèse.

$$\begin{aligned} P_2 &\equiv \sum_{i=1}^n \dot{w}_{ni}(x, h_n)(X_i - x) - 1 \\ &= \frac{1}{h_n S_n^2} \left\{ -S_n \sum_{i=1}^n \dot{K}\left(\frac{X_i - x}{h_n}\right)(X_i - x) \right. \\ &\quad \left. + \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right)(X_i - x) \sum_{j=1}^n \dot{K}\left(\frac{X_j - x}{h_n}\right) \right\} - 1. \end{aligned} \quad (2.2)$$

Avec le développement suivant, on peut réécrire $\dot{K}\left(\frac{X_i - x}{h_n}\right)(X_i - x)$:

$$\begin{aligned} \frac{d}{dx} \left[K\left(\frac{X_i - x}{h_n}\right)(X_i - x) \right] &= -\frac{1}{h_n} \dot{K}\left(\frac{X_i - x}{h_n}\right)(X_i - x) - K\left(\frac{X_i - x}{h_n}\right). \\ \Rightarrow \dot{K}\left(\frac{X_i - x}{h_n}\right) &= -h_n \left[\frac{d}{dx} \left(K\left(\frac{X_i - x}{h_n}\right)(X_i - x) \right) + K\left(\frac{X_i - x}{h_n}\right) \right]. \end{aligned}$$

On remplace $\dot{K}\left(\frac{X_i - x}{h_n}\right)(X_i - x)$ dans l'équation (2.2) et on utilise à nouveau la Proposition 1.1 :

$$\begin{aligned}
P_2 &= \frac{1}{h_n S_n^2} \left\{ h_n S_n \sum_{i=1}^n \left(\frac{d}{dx} \left(K \left(\frac{X_i - x}{h_n} \right) (X_i - x) \right) + K \left(\frac{X_i - x}{h_n} \right) \right) \right. \\
&\quad \left. + \sum_{i=1}^n K \left(\frac{X_i - x}{h_n} \right) \sum_{j=1}^n \dot{K} \left(\frac{X_j - x}{h_n} \right) \right\} - 1 \\
&= \frac{1}{h_n S_n^2} \left\{ h_n S_n \frac{d}{dx} \sum_{i=1}^n \left\{ K \left(\frac{X_i - x}{h_n} \right) (X_i - x) \right\} + h_n S_n \sum_{i=1}^n K \left(\frac{X_i - x}{h_n} \right) \right. \\
&\quad \left. + \sum_{i=1}^n K \left(\frac{X_i - x}{h_n} \right) (X_i - x) \sum_{j=1}^n \dot{K} \left(\frac{X_j - x}{h_n} \right) \right\} - 1 \\
&= \frac{1}{h_n S_n^2} \left\{ h_n S_n \frac{d}{dx} \sum_{i=1}^n \left\{ K \left(\frac{X_i - x}{h_n} \right) (X_i - x) \right\} \right. \\
&\quad \left. + \sum_{i=1}^n K \left(\frac{X_i - x}{h_n} \right) (X_i - x) \sum_{j=1}^n \dot{K} \left(\frac{X_j - x}{h_n} \right) \right\}.
\end{aligned}$$

On utilise le fait que $S_n = O_p(nh_n)$, $\sum_{i=1}^n K \left(\frac{X_i - x}{h_n} \right) = O_p(nh_n^3)$ et $\sum_{i=1}^n \dot{K} \left(\frac{X_i - x}{h_n} \right) = O_p(nh_n^2)$ pour obtenir que

$$P_2 = O_p(h_n^2).$$

Pour (P3),

$$\begin{aligned}
\sum_{i=1}^n \dot{w}_{ni}(x, h_n) (X_i - x)^2 &= \frac{1}{h_n S_n^2} \left\{ -S_n \sum_{i=1}^n \dot{K} \left(\frac{X_i - x}{h_n} \right) (X_i - x)^2 \right. \\
&\quad \left. + \sum_{i=1}^n K \left(\frac{X_i - x}{h_n} \right) (X_i - x)^2 \sum_{j=1}^n \dot{K} \left(\frac{X_j - x}{h_n} \right) \right\}.
\end{aligned}$$

On a déjà que

$$\sum_{i=1}^n K \left(\frac{X_i - x}{h_n} \right) (X_i - x)^2 = O_p(nh_n^3) \quad \text{et} \quad \sum_{i=1}^n \dot{K} \left(\frac{X_i - x}{h_n} \right) (X_i - x)^2 = O_p(nh_n^4),$$

on obtient :

$$\begin{aligned}
\sum_{i=1}^n \dot{w}_{ni}(x, h_n)(X_i - x)^2 &= \frac{1}{h_n S_n^2} \left\{ -S_n \sum_{i=1}^n \dot{K} \left(\frac{X_i - x}{h_n} \right) (X_i - x)^2 \right. \\
&\quad \left. + \sum_{i=1}^n K \left(\frac{X_i - x}{h_n} \right) (X_i - x)^2 \sum_{j=1}^n \dot{K} \left(\frac{X_j - x}{h_n} \right) \right\} \\
&= O_p(h_n^2).
\end{aligned}$$

On calcule ensuite une expression $\dot{w}_{ni}^2(x, h_n)$ pour vérifier la quatrième propriété :

$$\begin{aligned}
\dot{w}_{ni}^2(x, h_n) &= \frac{1}{h_n^2 S_n^4} \left[S_n^2 \left\{ \dot{K} \left(\frac{X_i - x}{h_n} \right) \right\}^2 \right. \\
&\quad \left. - 2K \left(\frac{X_i - x}{h_n} \right) \dot{K} \left(\frac{X_i - x}{h_n} \right) \sum_{j=1}^n \dot{K} \left(\frac{X_j - x}{h_n} \right) \right. \\
&\quad \left. + K^2 \left(\frac{X_i - x}{h_n} \right) \sum_{j=1}^n \sum_{l=1}^n \dot{K} \left(\frac{X_j - x}{h_n} \right) \dot{K} \left(\frac{X_l - x}{h_n} \right) \right].
\end{aligned}$$

Pour regarder le comportement de $\sum_{i=1}^n \dot{w}_{ni}^2(x, h_n)$, on utilise les deux résultats suivants :

$$\sum_{i=1}^n \left\{ \dot{K} \left(\frac{X_i - x}{h_n} \right) \right\}^2 = O(nh_n) \quad \text{et} \quad \sum_{i=1}^n K^2 \left(\frac{X_i - x}{h_n} \right) = O(nh_n),$$

ainsi que

$$2K \left(\frac{X_i - x}{h_n} \right) \dot{K} \left(\frac{X_i - x}{h_n} \right) = -h_n \frac{d}{dx} \left(K^2 \left(\frac{X_i - x}{h_n} \right) \right)$$

pour conclure :

$$\begin{aligned}
\sum_{i=1}^n \dot{w}_{ni}^2(x, h_n) &= \frac{1}{h_n^2 S_n^4} \sum_{i=1}^n \left[S_n^2 \left\{ \dot{K} \left(\frac{X_i - x}{h_n} \right) \right\}^2 \right. \\
&\quad \left. - 2K \left(\frac{X_i - x}{h_n} \right) \dot{K} \left(\frac{X_i - x}{h_n} \right) S_n \right. \\
&\quad \left. + K^2 \left(\frac{X_i - x}{h_n} \right) \sum_{j=1}^n \sum_{l=1}^n \dot{K} \left(\frac{X_j - x}{h_n} \right) \dot{K} \left(\frac{X_l - x}{h_n} \right) \right] \\
&= O_p \left(\frac{1}{nh_n^3} \right).
\end{aligned}$$

Pour démontrer (P5), on a : $\sum_{i=1}^n \left\{ \dot{K} \left(\frac{X_i - x}{h_n} \right) \right\}^2 (X_i - x) = O_p(nh_n^3)$,

$\sum_{i=1}^n K \left(\frac{X_i - x}{h_n} \right) \dot{K} \left(\frac{X_i - x}{h_n} \right) (X_i - x) = O_p(nh_n^2)$ et $\sum_{i=1}^n K^2 \left(\frac{X_i - x}{h_n} \right) = O_p(nh_n^3)$. On en conclut

que

$$\begin{aligned}
\sum_{i=1}^n \dot{w}_{ni}^2(x, h_n)(X_i - x) &= \frac{1}{h_n^2 S_n^4} \left\{ S_n^2 \sum_{i=1}^n \left\{ \dot{K} \left(\frac{X_i - x}{h_n} \right) \right\}^2 (X_i - x) \right. \\
&\quad - 2 \sum_{i=1}^n K \left(\frac{X_i - x}{h_n} \right) \dot{K} \left(\frac{X_i - x}{h_n} \right) (X_i - x) \sum_{j=1}^n \dot{K} \left(\frac{X_j - x}{h_n} \right) \\
&\quad \left. + \sum_{i=1}^n K^2 \left(\frac{X_i - x}{h_n} \right) (X_i - x) \sum_{j=1}^n \sum_{l=1}^n \dot{K} \left(\frac{X_j - x}{h_n} \right) \dot{K} \left(\frac{X_l - x}{h_n} \right) \right\} \\
&= O_p \left(\frac{1}{nh_n} \right).
\end{aligned}$$

On peut également vérifier que les conditions tiennent pour les poids locaux linéaires, mais cela demande des calculs plus élaborés.

Proposition 2.3 (Espérance et Variance de \dot{F}_{xh}). *Sous les conditions (H), (F) et (P), on a :*

1. $E(\dot{F}_{xh}(y)) = \dot{F}_x(y) + o(1),$
 2. $\text{Var} \left(\dot{F}_{xh}(y) \right) = \frac{C}{nh_n^3} F_x(y)(1 - F_x(y)) + o \left(\frac{1}{nh_n^3} \right),$
- $C = \lim_{n \rightarrow \infty} nh_n^3 \sum_{i=1}^n \dot{w}_{ni}^2(x, h_n).$

Démonstration.

$$\begin{aligned}
E \left(\dot{F}_{xh}(y) \right) &= E \left(\sum_{i=1}^n \mathbb{1}(Y_i \leq y) \dot{w}_{ni}(x, h_n) \right) \\
&= \sum_{i=1}^n F_{x_i}(y) \dot{w}_{ni}(x, h_n)
\end{aligned}$$

On fait un développement de Taylor de F_{x_i} autour de x

$$\begin{aligned}
E(\dot{F}_{xh}(y)) &= F_x(y) \sum_{i=1}^n \dot{w}_{ni}(x, h_n) + \dot{F}_x(y) \sum_{i=1}^n \dot{w}_{ni}(x, h_n)(x_i - x) \\
&\quad + \frac{1}{2} \sum_{i=1}^n \ddot{F}_{z_i}(y) \dot{w}_{ni}(x, h_n)(x_i - x)^2.
\end{aligned}$$

où z_i est dans un voisinage de x_i .

D'après (F2) et (P3), on a

$$E(\dot{F}_{xh}(y)) = F_x(y) \sum_{i=1}^n \dot{w}_{ni}(x, h_n) + \dot{F}_x(y) \sum_{i=1}^n \dot{w}_{ni}(x, h_n)(x_i - x) + o(1).$$

À l'aide de (P1)-(P2), on conclut que $E(\dot{F}_{xh}(y)) = \dot{F}_x(y) + o(1)$.

Calculons la variance de \dot{F}_{xh} . D'après (P5),

$$\text{Var} \left(\dot{F}_{xh}(y) \right) = F_x(y)(1 - F_x(y)) \sum_{i=1}^n \dot{w}_{ni}^2(x, h_n) + o \left(\frac{1}{nh_n^3} \right). \quad (2.3)$$

En posant $C = \lim_{n \rightarrow \infty} nh_n^3 \sum_{i=1}^n \dot{w}_{ni}^2(x, h_n)$ et en utilisant (P4), on obtient que

$$\text{Var} \left(\dot{F}_{xh}(y) \right) = \frac{C}{nh_n^3} F_x(y)(1 - F_x(y)) + o \left(\frac{1}{nh_n^3} \right).$$

□

Le lemme suivant établit le biais de l'estimateur de la dérivée de la distribution conditionnelle pour les poids de Nadaraya-Watson.

Lemme 2.1.1 (Biais de \dot{F}_{xh} avec NW). Sous les conditions (H), (F) et (P) et avec les poids de Nadaraya-Watson, on a que

$$E(\dot{F}_{xh}(y)) = \dot{F}_x(y) + h_n^2 \left[C_1 \dot{F}_x(y) + C_2 \frac{\ddot{F}_x(y)}{2} \right] + o(h_n^2),$$

où $C_1 = \lim_{n \rightarrow \infty} \frac{1}{h_n^2} (\sum_{i=1}^n \dot{w}_{ni}(x, h_n)(x_i - x) - 1)$ et $C_2 = \lim_{n \rightarrow \infty} \frac{1}{h_n^2} \sum_{i=1}^n \dot{w}_{ni}(x, h_n)(x_i - x)^2$.

La prochaine proposition donne le paramètre de lissage optimal minimisant l'AMISE de l'estimateur de la dérivée de la distribution conditionnelle.

Proposition 2.4. *Sous les conditions (F) et (P) et avec les poids de Nadaraya-Watson, on a que l'erreur quadratique asymptotique moyenne intégrée de \dot{F}_{xh} est donnée par*

$$AMISE(\dot{F}_{xh}) = \frac{1}{nh_n^3} \alpha + h_n^4 \beta.$$

où $\alpha = \int_{\text{supp}(Y)} CF_x(y)(1 - F_x(y))dy$ et $\beta = \int_{\text{supp}(Y)} [C_1\dot{F}_x(y) + C_2\frac{\ddot{F}_x(y)}{2}]^2dy$. Le paramètre h_n optimal minimisant l'AMISE de \dot{F}_{xh} est

$$h^* = \left(\frac{3\alpha}{4\beta}\right)^{\frac{1}{7}} n^{-\frac{1}{7}}. \quad (2.4)$$

En considérant cette valeur de h_n , on obtient l'AMISE optimale :

$$AMISE^*(\dot{F}_{xh}) = An^{-\frac{4}{7}} \quad \text{où } A = \left(\left(\frac{4}{3}\right)^{\frac{3}{7}} + \left(\frac{3}{4}\right)^{\frac{4}{7}}\right) \alpha^{\frac{4}{7}} \beta^{\frac{3}{7}}.$$

Le prochain résultat établit la convergence en probabilité de \dot{F}_{xh} .

Proposition 2.5 (Convergence en probabilité de \dot{F}_{xh}). *Sous les conditions (H), (F) et (P),*

$$\dot{F}_{xh}(y) \xrightarrow{P} \dot{F}_x(y) \quad \text{quand } n \rightarrow \infty.$$

Démonstration. On utilise l'inégalité de Chebyshev (voir [CB01]). Soit $\varepsilon > 0$, alors

$$\begin{aligned} P(|\dot{F}_{xh}(y) - E(\dot{F}_x(y))| > \varepsilon) &\leq \frac{1}{\varepsilon^2} \text{Var}(\dot{F}_{xh}) \\ &= \frac{1}{\varepsilon^2} \sum_{i=1}^n \dot{w}_{ni}^2(x, h_n) F_{x_i}(y)(1 - F_{x_i}(y)) \\ &\leq \frac{1}{\varepsilon^2} \sum_{i=1}^n \dot{w}_{ni}^2(x, h_n) \\ &= O\left(\frac{1}{nh_n^3}\right) \quad \text{d'après (P4).} \end{aligned}$$

$$\Rightarrow \lim_{n \rightarrow \infty} P(|\dot{F}_{xh}(y) - E(\dot{F}_x(y))| > \varepsilon) = 0.$$

En utilisant les conditions (F) et (P1)-(P3), on a que

$$|E(\dot{F}_{xh}(y)) - \dot{F}_x(y)| \rightarrow 0 \quad \text{quand } n \rightarrow \infty.$$

À l'aide de l'inégalité triangulaire :

$$|\dot{F}_{xh}(y) - \dot{F}_x(y)| \leq |\dot{F}_{xh}(y) - E(\dot{F}_{xh}(y))| + |E(\dot{F}_{xh}(y)) - \dot{F}_x(y)|, \quad (2.5)$$

on en conclut que $\dot{F}_{xh}(y) \xrightarrow{P} \dot{F}_x(y)$. □

Le prochain exemple aborde l'évaluation de la MISE de l'estimateur de la dérivée de la fonction de répartition conditionnelle pour différentes tailles d'échantillons.

Exemple 2.1. Soit $X \sim U[0, 1]$ et $Y|X \sim \exp(\theta(X))$ où $\theta(z) = 0.5 - 4z^2 + 4z$. On peut en déduire que

$$\dot{F}_x(y) = -\frac{y\theta'(x)}{\theta^2(x)}e^{-\frac{y}{\theta(x)}}.$$

On simule 200 échantillons de taille n et on approxime ensuite l'erreur quadratique intégrée par

$$\widehat{MISE}(\dot{F}_{xh}) = \frac{1}{200} \sum_{b=1}^{200} \sum_{i=1}^n [\dot{F}_{xh}^{(b)}(y_i) - \dot{F}_x(y_i)]^2,$$

où $h = n^{-\frac{1}{7}}$. On répète l'expérience pour plusieurs taille d'échantillon et on retrouve les résultats à la Figure 2.1. On constate que l'erreur quadratique intégrée de \dot{F}_{xh} ne converge

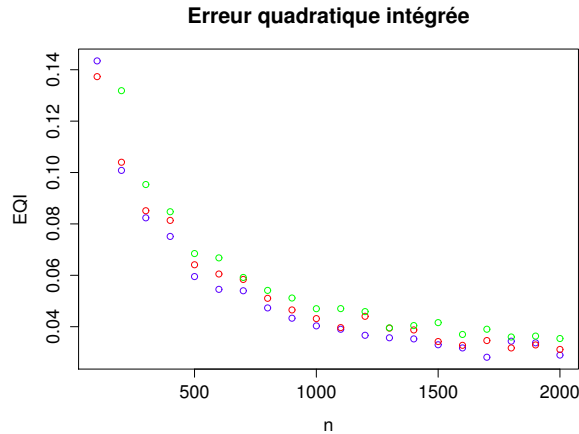


FIGURE 2.1 – Approximation de l'erreur quadratique intégrée de l'estimateur de la dérivée de la distribution conditionnelle pour plusieurs tailles d'échantillon pour le modèle donné dans l'Exemple 2.1. L'expérience est faite pour trois valeurs de la covariable X ($x = 0.5$ (bleu), $x = 0.6$ (rouge) et $x = 0.7$ (vert)).

pas rapidement vers 0 en tant que fonction de n . On a besoin d'une taille d'échantillon très grande pour avoir une bonne précision de l'estimateur.

2.2 Convergence en loi

Pour montrer la convergence en loi vers une variable aléatoire distribuée selon une normale, on utiliserait le théorème central limite. Cependant, dans un design fixe, les variables aléatoires $\{(Y_i, X_i)\}_{i=1}^n$ ne sont pas identiquement distribuées, on doit donc utiliser une version moderne de Lyapunov du théorème central limite (voir [vdV00]).

Lemme 2.2.1 (Théorème central limite de Lyapunov). Soit V_1, V_2, \dots, V_n , une suite de variables aléatoires telles que $E(V_i) = \mu_i$, $s_n^2 = \sum_{i=1}^n \text{Var}(V_i)$ et $\delta > 0$. Si

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^n E\{|V_i - \mu_i|^{2+\delta}\} = 0,$$

alors

$$\frac{1}{s_n} \sum_{i=1}^n (V_i - \mu_i) \xrightarrow{d} N(0, 1).$$

Proposition 2.6 (Convergence en loi de \dot{F}_{xh} avec NW). *Sous les mêmes conditions que la Proposition 2.5 ainsi qu'avec la condition :*

$$(E3) \quad \sum_{i=1}^n \dot{w}_{ni}^4(x, h_n) = O\left(\frac{1}{n^3 h_n^7}\right),$$

on a avec les poids de Nadaraya-Watson que

$$\sqrt{nh_n^3} \left[\dot{F}_{xh}(y) - \dot{F}_x(y) \right] \xrightarrow{d} N(\bar{B}_x(y), C(F_x(y)(1 - F_x(y))),$$

où $\bar{B}_x(y) = C_1 \dot{F}_x(y) + C_2 \ddot{F}_x(y)$ et $C = \lim_{n \rightarrow \infty} nh_n^3 \sum_{i=1}^n \dot{w}_{ni}^2(x, h_n)$.

Démonstration. On pose $V_i = \sqrt{nh_n^3} \mathbf{1}(Y_i \leq y) \dot{w}_{ni}(x, h_n)$. On a donc que $\mu_i = E(\sqrt{nh_n^3} \mathbf{1}(Y_i \leq y) \dot{w}_{ni}(x, h_n)) = \sqrt{nh_n^3} \dot{w}_{ni}(x, h_n) F_{x_i}(y)$. On peut prendre $\delta = 2$ et on va montrer que

$$\frac{1}{s_n^4} \sum_{i=1}^n E\{(V_i - \mu_i)^4\} \longrightarrow 0.$$

Or, on a que

$$\begin{aligned} \mathbb{E}\{(V_i - \mu_i)^4\} &= n^2 h_n^6 \dot{w}_{ni}^4(x, h_n) \mathbb{E}\{(\mathbb{1}(Y_i \leq y) - F_{x_i}(y))^4\} \\ &\leq n^2 h_n^6 \dot{w}_{ni}^4(x, h_n). \end{aligned}$$

À l'aide de la condition (E3), on en conclut que

$$\sum_{i=1}^n \mathbb{E}\{(V_i - \mu_i)^4\} = O(n^2 h_n^6) \sum_{i=1}^n \dot{w}_{ni}^4(x, h_n) = O(n^2 h_n^6) O\left(\frac{1}{n^3 h_n^7}\right) = O\left(\frac{1}{n h_n}\right).$$

On regarde maintenant le terme s_n^2 .

$$\begin{aligned} s_n^2 &= \sum_{i=1}^n \text{Var}(V_i) \\ &= \sum_{i=1}^n \text{Var}(\sqrt{n h_n^3} \mathbb{1}(Y_i \leq y) \dot{w}_{ni}(x, h_n)) \\ &= n h_n^3 \sum_{i=1}^n \dot{w}_{ni}^2(x, h_n) \text{Var}(\mathbb{1}(Y_i \leq y)) \\ &= n h_n^3 \sum_{i=1}^n \dot{w}_{ni}^2(x, h_n) F_{x_i}(y) (1 - F_{x_i}(y)) \\ &= n h_n^3 \left[\sum_{i=1}^n \dot{w}_{ni}^2(x, h_n) [F_x(y) (1 - F_x(y)) + o\left(\frac{1}{n h_n^3}\right)] \right] \\ &= C F_x(y) (1 - F_x(y)) + o(1) \\ &\Rightarrow \frac{1}{s_n^4} \sum_{i=1}^n \mathbb{E}\{(V_i - \mu_i)^4\} = O\left(\frac{1}{n h_n}\right). \end{aligned}$$

Par conséquent, la condition est vérifiée et on a bien que

$$\frac{\sqrt{n h_n^3}}{\text{Var}(F_{xh}(y))} (\dot{F}_{xh}(y) - \mathbb{E}(\dot{F}_{xh}(y))) \xrightarrow{L} N(0, 1).$$

Avec les poids de Nadaraya-Watson,

$$\begin{aligned}
\sqrt{nh_n^3}(\mathbb{E}(\dot{F}_{xh}(y)) - \dot{F}_x(y)) &= \sqrt{nh_n^3} \left\{ \sum_{i=1}^n \dot{w}_{ni}(x, h_n) F_{x_i}(y) - \dot{F}_x(y) \right\} \\
&= \sqrt{nh_n^3} \sum_{i=1}^n \dot{w}_{ni}(x, h_n) \left\{ F_x(y) + \dot{F}_x(y)(x_i - x) \right. \\
&\quad \left. + \ddot{F}_{z_i}(y) \frac{(x_i - x)^2}{2} \right\} - \sqrt{nh_n^3} \dot{F}_x(y) \\
&= \sqrt{nh_n^3} \dot{F}_x(y) \left\{ \sum_{i=1}^n \dot{w}_{ni}(x, h_n)(x_i - x) - 1 \right\} \\
&\quad + \sqrt{nh_n^3} \left\{ \sum_{i=1}^n \frac{\ddot{F}_{z_i}(y)}{2} \dot{w}_{ni}(x, h_n)(x_i - x)^2 \right\} \\
&= C_1 \dot{F}_x(y) + C_2 \frac{\ddot{F}_x(y)}{2} + o(1).
\end{aligned}$$

On obtient alors

$$\sqrt{nh_n^3} \{ \dot{F}_{xh}(y) - \mathbb{E}(\dot{F}_{xh}(y)) + \mathbb{E}(\dot{F}_{xh}(y)) - \dot{F}_x(y) \} \xrightarrow{L} N(\bar{B}_x(y), CF_x(y)(1 - F_x(y))),$$

$$\text{où } \bar{B}_x(y) = C_1 \dot{F}_x(y) + C_2 \frac{\ddot{F}_x(y)}{2}.$$

□

Dans l'optique d'estimer le biais de C_{xh} lorsque $C_x(u, v) = uv$, l'inégalité suivante peut nous permettre de nous débarrasser de la double structure aléatoire.

$$\begin{aligned}
|\dot{F}_{1xh}\{F_{1xh}^{-1}(u)\} - \dot{F}_{1x}\{F_{1x}^{-1}(u)\}| &\leq |\dot{F}_{1xh}\{F_{1xh}^{-1}(u)\} - \dot{F}_{1x}\{F_{1xh}^{-1}(u)\}| \\
&\quad + |\dot{F}_{1x}\{F_{1xh}^{-1}(u)\} - \dot{F}_{1x}\{F_{1x}^{-1}(u)\}| \\
&\leq \sup_{y \in \mathbb{R}} |\dot{F}_{1xh}(y) - \dot{F}_x(y)| + D
\end{aligned}$$

$$\text{où } D = |\dot{F}_{1x}\{F_{1xh}^{-1}(u)\} - \dot{F}_{1x}\{F_{1x}^{-1}(u)\}|.$$

Pour obtenir la convergence en probabilité de $\tilde{B}_{H_{xh}}$ vers \tilde{B}_{H_x} , il suffit de démontrer la convergence uniforme en probabilité de \dot{F}_{xh} vers \dot{F}_x et prouver que le terme D est négligeable.

[VOG11] ont démontré que l'estimateur du quantile conditionnel noté F_{1xh}^{-1} converge en probabilité vers F_{1x}^{-1} . De plus, en utilisant la condition (F1), on peut appliquer le théorème des applications continues et on obtient que

$$D = |\dot{F}_{1x}\{F_{1xh}^{-1}(u)\} - \dot{F}_{1x}\{F_{1x}^{-1}(u)\}| \xrightarrow{P} 0.$$

On passe ensuite à la convergence uniforme en probabilité de \dot{F}_{xh} .

2.3 Convergence uniforme en probabilité de \dot{F}_{xh}

Pour démontrer la convergence uniforme en probabilité de \dot{F}_{xh} , on a besoin de nouvelles hypothèses concernant le système de poids.

Hypothèses (E)

- (E1) $\sum_{i=1}^n |\dot{w}_{ni}(x, h_n)| = O\left(\frac{1}{h_n}\right),$
- (E2) $\sum_{i=1}^n |\dot{w}_{ni}(x, h_n)| |x_i - x| = O(1),$
- (E3) $\sum_{i=1}^n |\dot{w}_{ni}(x, h_n)| |x_i - x|^2 = O(h_n).$

Proposition 2.7. *Les poids de Nadaraya-Watson vérifient les conditions (E).*

Démonstration. La démonstration est basée sur l'équation (2.1) et l'inégalité triangulaire :

$$\begin{aligned} |\dot{w}_{ni}(x, h_n)| &= \left| \frac{1}{h_n S_n^2} \left[-S_n \dot{K} \left(\frac{X_i - x}{h_n} \right) + K \left(\frac{X_i - x}{h_n} \right) \sum_{j=1}^n \dot{K} \left(\frac{X_j - x}{h_n} \right) \right] \right| \\ &\leq \frac{1}{h_n S_n^2} \left[\left| S_n \dot{K} \left(\frac{X_i - x}{h_n} \right) \right| + \left| K \left(\frac{X_i - x}{h_n} \right) \sum_{j=1}^n \dot{K} \left(\frac{X_j - x}{h_n} \right) \right| \right] \\ &\leq \frac{1}{h_n S_n^2} \left[S_n \left| \dot{K} \left(\frac{X_i - x}{h_n} \right) \right| + K \left(\frac{X_i - x}{h_n} \right) \sum_{j=1}^n \left| \dot{K} \left(\frac{X_j - x}{h_n} \right) \right| \right], \end{aligned}$$

$$\Rightarrow \sum_{i=1}^n |\dot{w}_{ni}(x, h_n)| \leq \frac{1}{h_n S_n^2} \left[S_n \sum_{i=1}^n \left| \dot{K} \left(\frac{X_i - x}{h_n} \right) \right| + \sum_{j=1}^n K \left(\frac{X_i - x}{h_n} \right) \sum_{j=1}^n \left| \dot{K} \left(\frac{X_j - x}{h_n} \right) \right| \right].$$

Dû au fait que $\int_{-1}^1 |\dot{K}(v)| dv \neq 0$, on peut montrer que

$$\sum_{i=1}^n \left| \dot{K} \left(\frac{X_i - x}{h_n} \right) \right| = O_p(nh_n).$$

On obtient alors que

$$\sum_{i=1}^n |\dot{w}_{ni}(x, h_n)| = O_p \left(\frac{1}{h_n} \right). \quad (2.6)$$

Pour (E2), on a que

$$\begin{aligned} \sum_{i=1}^n |\dot{w}_{ni}(x, h_n)| |X_i - x| &\leq \frac{1}{h_n S_n^2} \left\{ S_n \sum_{i=1}^n \left| \dot{K} \left(\frac{X_i - x}{h_n} \right) \right| |X_i - x| \right. \\ &\quad \left. + \sum_{i=1}^n K \left(\frac{X_i - x}{h_n} \right) |X_i - x| \sum_{j=1}^n \left| \dot{K} \left(\frac{X_j - x}{h_n} \right) \right| \right\}. \end{aligned}$$

On a besoin de développer $\sum_{i=1}^n \left| \dot{K} \left(\frac{X_i - x}{h_n} \right) \right| |X_i - x|$ et $\sum_{i=1}^n K \left(\frac{X_i - x}{h_n} \right) |X_i - x|$. Or, comme $\int_{-1}^1 \dot{K}(v) |v| dv > 0$ et $\int_{-1}^1 K(v) |v| dv > 0$, on en déduit que $\sum_{i=1}^n \left| \dot{K} \left(\frac{X_i - x}{h_n} \right) \right| |X_i - x| = O(nh_n^2)$ et $\sum_{i=1}^n K \left(\frac{X_i - x}{h_n} \right) |X_i - x| = O(nh_n^2)$. On en conclut que $\sum_{i=1}^n |\dot{w}_{ni}(x, h_n)| |X_i - x| = O_p(1)$.

Pour la démonstration de (E3), on a l'inégalité suivante :

$$\begin{aligned} \sum_{i=1}^n |\dot{w}_{ni}(x, h_n)| |X_i - x|^2 &\leq \frac{1}{h_n S_n^2} \left\{ S_n \sum_{i=1}^n \left| \dot{K} \left(\frac{X_i - x}{h_n} \right) \right| |X_i - x|^2 \right. \\ &\quad \left. + \sum_{i=1}^n K \left(\frac{X_i - x}{h_n} \right) |X_i - x|^2 \sum_{j=1}^n \left| \dot{K} \left(\frac{X_j - x}{h_n} \right) \right| \right\} \end{aligned}$$

Suivant le même raisonnement que précédemment, on montre que

$$\sum_{i=1}^n \left| \dot{K} \left(\frac{X_i - x}{h_n} \right) \right| |X_i - x|^2 = O_p(nh_n^3).$$

On utilise ensuite la Proposition 1.1 pour obtenir :

$$\sum_{i=1}^n |\dot{w}_{ni}(x, h_n)| |X_i - x|^2 = O_p(h_n). \quad (2.7)$$

□

Proposition 2.8 (Convergence uniforme en probabilité de \dot{F}_{xh}). *Sous les conditions (H), (F), (P) et (E), on a que :*

$$\sup_{y \in \mathbb{R}} |\dot{F}_{xh}(y) - \dot{F}_x(y)| \xrightarrow{P} 0.$$

Démonstration. Il est en général difficile d'évaluer une probabilité sur le supremum d'un ensemble infini. On construit judicieusement une grille composée d'un nombre fini d'éléments pour recouvrir \mathbb{R} . Soit δ tel que $0 < \delta < \frac{5}{3}$. On construit $\{\tilde{y}_i\}_{i=1}^m$ telle que $\tilde{y}_i = F_x^{-1}\left(\frac{i}{\sqrt{nh_n}^{\delta+1}}\right) \quad \forall i = 0 \dots [\sqrt{nh_n}^{\delta+1}]$. Avec cette grille, on obtient la relation suivante :

$$F_x(\tilde{y}_i) - F_x(\tilde{y}_{i-1}) \leq \frac{1}{\sqrt{nh_n}^{\delta+1}}. \quad (2.8)$$

Prenons maintenant en considération l'inégalité suivante :

$$\sup_{y \in \mathbb{R}} |\dot{F}_{xh}(y) - \dot{F}_x(y)| < \sup_{y \in \mathbb{R}} |\dot{F}_{xh}(y) - \mathbb{E}(\dot{F}_{xh}(y))| + \sup_{y \in \mathbb{R}} |\mathbb{E}(\dot{F}_{xh}(y)) - \dot{F}_x(y)|.$$

Soit $\mathbb{F}_{xh}(y) = \dot{F}_{xh}(y) - \mathbb{E}(\dot{F}_{xh}(y))$ et prenons \bar{y} et \underline{y} , les deux éléments de la grille tels que $\underline{y} < y < \bar{y}$. Alors on peut borner $\mathbb{F}_{xh}(y)$ par une quantité dépendant seulement des éléments de

la grille :

$$\begin{aligned}
\mathbb{F}_{xh}(y) &= \dot{F}_{xh}(y) - \mathbb{E}(\dot{F}_{xh}(y)) \\
&= \sum_{i=1}^n \{\mathbb{1}(Y_i \leq y) - F_{x_i}(y)\} \dot{w}_{ni}(x, h_n) \\
&\leq \sum_{i=1}^n |\mathbb{1}(Y_i \leq y) - F_{x_i}(y)| |\dot{w}_{ni}(x, h_n)| \\
&\leq \sum_{i=1}^n |\mathbb{1}(Y_i \leq \bar{y}) - F_{x_i}(\bar{y})| |\dot{w}_{ni}(x, h_n)| + \sum_{i=1}^n |F_{x_i}(\bar{y}) - \mathbb{1}(Y_i \leq \underline{y})| |\dot{w}_{ni}(x, h_n)| \\
&\leq \sum_{i=1}^n |\mathbb{1}(Y_i \leq \bar{y}) - F_{x_i}(\bar{y})| |\dot{w}_{ni}(x, h_n)| + \sum_{i=1}^n |\mathbb{1}(Y_i \leq \underline{y}) - F_{x_i}(\underline{y})| |\dot{w}_{ni}(x, h_n)| \\
&\quad + 2 \sum_{i=1}^n |F_{x_i}(\bar{y}) - F_{x_i}(\underline{y})| |\dot{w}_{ni}(x, h_n)| \\
&\leq 2 \max_{j \in \{1, \dots, m\}} \sum_{i=1}^n |\mathbb{1}(Y_i \leq \tilde{y}_j) - F_{x_i}(\tilde{y}_j)| |\dot{w}_{ni}(x, h_n)| + 2D^*,
\end{aligned}$$

où $D^* = \sum_{i=1}^n |F_{x_i}(\bar{y}) - F_{x_i}(\underline{y})| |\dot{w}_{ni}(x, h_n)|$. Montrons que D^* est un terme négligeable qui tend vers 0 quand $n \rightarrow \infty$. Pour ce faire, on fait un développement de Taylor :

$$\begin{aligned}
D^* &= |F_x(\bar{y}) - F_x(\underline{y})| \sum_{i=1}^n |\dot{w}_{ni}(x, h_n)| \\
&\quad + |\dot{F}_x(\bar{y}) - \dot{F}_x(\underline{y})| \sum_{i=1}^n |\dot{w}_{ni}(x, h_n)| |x_i - x| \\
&\quad + \sum_{i=1}^n |\ddot{F}_{z_i}(\bar{y}) - \ddot{F}_{z_i}(\underline{y})| |\dot{w}_{ni}(x, h_n)| \frac{|x_i - x|^2}{2},
\end{aligned}$$

où $z_i \in [x_i, x]$. D'après les hypothèses, on conclut que $D^* = o(1)$.

La prochaine étape consiste à montrer que $\sum_{i=1}^n |\mathbb{1}(Y_i \leq \tilde{y}_j) - F_{x_i}(\tilde{y}_j)| |\dot{w}_{ni}(x, h_n)|$ converge vers 0 en probabilité :

$$\begin{aligned}
P\left(\sum_{i=1}^n |\mathbb{1}(Y_i \leq \tilde{y}_j) - F_{x_i}(\tilde{y}_j)| |\dot{w}_{ni}(x, h_n)| > \epsilon\right) \\
&< \frac{1}{\epsilon^2} \mathbb{E} \left[\left\{ \sum_{i=1}^n |\mathbb{1}(Y_i \leq \tilde{y}_j) - F_{x_i}(\tilde{y}_j)| |\dot{w}_{ni}(x, h_n)| \right\}^2 \right] \\
&< \frac{1}{\epsilon^2} \left\{ \sum_{i=1}^n |\dot{w}_{ni}^2(x, h_n)| \right\}^2 \\
&= O\left(\frac{1}{n^2 h_n^6}\right).
\end{aligned}$$

Pour $\epsilon > 0$, on obtient que

$$\begin{aligned}
P\left(\max_{j \in \{1, \dots, m\}} \sum_{i=1}^n |\mathbb{1}(Y_i \leq \tilde{y}_j) - F_{x_i}(\tilde{y}_j)| |\dot{w}_{ni}(x, h_n)| > \epsilon\right) &\leq \sum_{j=1}^m P\left(\sum_{i=1}^n |\mathbb{1}(Y_i \leq \tilde{y}_j) - F_{x_i}(\tilde{y}_j)| |\dot{w}_{ni}(x, h_n)| > \epsilon\right) \\
&= O\left(\frac{m}{n^2 h_n^6}\right)
\end{aligned}$$

Comme $m = \lceil \sqrt{nh_n}^{\delta+1} \rceil$, $h_n = O(n^{-\frac{1}{7}})$ et que $\delta < \frac{5}{3}$, on démontre que M est aussi un terme négligeable. On a maintenant que

$$\sup_{y \in \mathbb{R}} |\dot{F}_{xh}(y) - \mathbb{E}(\dot{F}_{xh}(y))| \xrightarrow{P} 0.$$

Pour ce qui est du terme $\sup_{y \in \mathbb{R}} |\mathbb{E}(\dot{F}_{xh}(y)) - \dot{F}_x(y)|$, on a déjà démontré dans la Proposition 2.3 que cette expression allait vers 0 quand $n \rightarrow \infty$. Ceci conclut la démonstration. \square

Avec ce résultat, on est assuré que le biais de H_x sous H_0 sera bien estimé par $\tilde{B}_{H_{xh}}$ comme on a que

$$\tilde{B}_{H_{xh}} \xrightarrow{P} \tilde{B}_{H_x}.$$

Maintenant qu'on a établi les propriétés de \dot{F}_{xh} , on est prêt à discuter du test d'indépendance basé sur le processus \mathbb{C}_{xh} donné dans l'équation (1.21).

Chapitre 3

Tests d'indépendance conditionnelle et simulations

Dans ce chapitre, on se concentre sur un test d'indépendance basé sur la copule conditionnelle.

3.1 Construction du test

Comme la dépendance conditionnelle de $(Y_1, Y_2)|X = x$ est totalement capturée par la copule C_x , on peut construire un test d'indépendance suivant les mêmes idées que [Deh81] pour la copule conditionnelle. Celui-ci consiste à tester

$$H_0 : C_x = C_0 \quad \text{contre} \quad H_1 : C_x \neq C_0.$$

Pour faire le test, on doit choisir une statistique de test qui nous permet d'évaluer l'écart entre un estimateur de C_x et C_0 . Le choix le plus commun est la fonctionnelle de Cramér-

von Mises définie par

$$S(f) = \int_0^1 \int_0^1 f^2(u, v) du dv \quad \text{pour } f : [0, 1]^2 \rightarrow \mathbb{R}.$$

On applique cette statistique au processus suivant :

$$\mathbb{C}_{xh}^0 = \sqrt{nh_n}(C_{xh} - C_0).$$

On rejette l'hypothèse H_0 si $S(\mathbb{C}_{xh}^0)$ dépasse un certain seuil. Ce seuil correspond normalement au quantile d'ordre $1-\alpha$ de la distribution asymptotique de la statistique de test $S(\mathbb{C}_{xh}^0)$ sous H_0 .

En pratique, celle-ci est inconnue et on doit l'approximer. On sait par contre que \mathbb{C}_{xh}^0 converge faiblement vers un élément de $l^\infty([0, 1])$:

$$\mathbb{C}_{xh}^0 \rightsquigarrow \mathbb{G}_{xh}^0 = V_{x0}\mathbb{G}_0 + E_{x0}\tilde{B}_{H_x}.$$

Une application directe du théorème des applications continues nous permet de montrer que

$$S(\mathbb{C}_{xh}^0) \rightsquigarrow S(\mathbb{G}_{xh}^0).$$

Si $\{\mathbb{G}_{xh}^{(b)}\}_{b=1}^B$ sont B copies valides de \mathbb{G}_{xh}^0 et qu'on pose $S^{(b)} = \int_0^1 \int_0^1 \mathbb{G}_{xh}^{(b)}(u, v) du dv$, alors l'ensemble $\{S^{(b)}\}_{b=1}^B$ forme un échantillon valide de la statistique $S(\mathbb{C}_{xh}^0)$. Pour faciliter la lecture, on décide de poser $S_{xh} = S(\mathbb{C}_{xh}^0)$.

Le défi maintenant est de développer une technique pour obtenir les répliques de \mathbb{G}_{xh}^0 . La prochaine section explique en détail notre méthode utilisant l'estimateur donné dans le chapitre trois pour estimer le biais de H_{xh} sous H_0 . On explique ensuite comment calculer facilement la statistique de Cramér-von Mises pour chaque copie du processus. Pour finir, des simulations sont faites pour comparer le niveau et la puissance du test.

3.2 Réplication de \mathbb{G}_{xh}^0

Par des résultats portant sur l'estimateur de la fonction de répartition bidimensionnelle, on sait que l'on peut répliquer facilement un pont Brownien bidimensionnelle en utilisant la technique suivante :

1. On génère $\{U_{1i}^{(b)}\}_{i=1}^n \sim U(0, 1)$ et $\{U_{2i}^{(b)}\}_{i=1}^n \sim U(0, 1)$ indépendantes.
2. On pose

$$\mathbb{B}_n^{(b)}(u, v) = \sqrt{n} \{H_n^{(b)}(u, v) - uv\} = n^{-\frac{1}{2}} \sum_{i=1}^n \{\mathbb{1}(U_{1i} \leq u, U_{2i} \leq v) - uv\},$$

où H_n est la distribution empirique de l'échantillon $\{(U_{1i}^{(b)}, U_{2i}^{(b)})\}_{i=1}^n$.

3. $\mathbb{B}_n^{(b)}(u, v)$ est une réplication valide de \mathbb{B} .

Le théorème des applications continues nous permet de confirmer que

$$\mathbb{C}_n^{(b)} = \mathbb{B}_n^{(b)}(u, v) - u\mathbb{B}_n^{(b)}(1, v) - v\mathbb{B}_n^{(b)}(u, 1) \rightsquigarrow \mathbb{G}_0.$$

Soit $\bar{\sigma}$, l'étendue interquartile de l'échantillon $\{X_i\}_{i=1}^n$. On peut donc répliquer \mathbb{G}_{xh}^0 en utilisant l'algorithme suivant. $\forall b = 1 \dots M$

1. Générer $(U_{1i}^{(b)}, U_{2i}^{(b)}) \sim C_0(u, v) = uv \quad \forall i = 1 \dots n$.
2. Poser $\mathbb{C}_n^{(b)} = \mathbb{B}_n^{(b)}(u, v) - u\mathbb{B}_n^{(b)}(1, v) - v\mathbb{B}_n^{(b)}(u, 1)$.
3. Utiliser le paramètre de lissage $h_n = \bar{\sigma}n^{-\frac{1}{5}}$ pour calculer C_{xh} ,
4. Calculer $\tilde{B}_{H_{xg}}$ avec le paramètre de lissage $g = \bar{\sigma}n^{-\frac{1}{7}}$.
5. $\mathbb{G}_{xh}^{(b)} = V_{xh}\mathbb{C}_n^{(b)} + E_{xh}\tilde{B}_{H_{xg}}$ est une réplication valide de \mathbb{G}_{xh}^0 .

Soit $\{\mathbb{G}_{xh}^{(b)}\}_{i=1}^n$, B indépendantes copies de \mathbb{C}_{xh}^0 . On ne peut pas appliquer la statistique de Cramér-von Mises sur chaque réplication, car l'expression du biais $\tilde{B}_{H_{xh}}$ est trop difficile à intégrer. On peut approximer la double intégrale en calculant une moyenne pour une

grille fixe donnée $\{g_i\}_{i=1}^m$ partitionnant l'intervalle $[0, 1]$. On obtient une approximation des statistiques pour chaque réplication :

$$\widehat{S}^{(b)} = \frac{1}{m^2} \sum_{l=1}^m \sum_{k=1}^m \mathbb{G}_{xh}^{(b)}(g_k, g_l).$$

Pour calculer la zone de rejet du test, on calcule le quantile d'ordre $(1 - \alpha)$ de l'ensemble $\{\widehat{S}^{(b)}\}$ et on le note K_α . On rejettera H_0 lorsque $S_{xh} > K_\alpha$.

On doit maintenant valider le test d'indépendance en vérifiant ses capacités à garder le niveau et à être puissant.

3.3 Validité du test et simulations

Fixons α , le seuil de tolérance. Supposons que l'on décide de simuler N jeux de données à partir d'un modèle sous H_0 et que l'on décide d'appliquer notre test pour calculer la proportion de fois où on a rejeté H_0 . Cette proportion constitue le niveau expérimental du test. Un test d'hypothèses valide doit conserver le niveau α .

Si au contraire, on décide de simuler N jeux de données sous H_1 , on souhaite que cette proportion soit égale à 1. Dans ce cas, on parlera de la puissance du test.

Pour les simulations, on considère deux modèles différents utilisant deux copules connues. La première copule est celle de Gumbel définie par

$$C_{gu}(u, v) = \exp \left(- \left((-\log(u))^\theta + (-\log(v))^\theta \right)^{\frac{1}{\theta}} \right), \quad \theta \geq 1.$$

On obtient la copule d'indépendance quand $\theta = 1$. La deuxième est la copule gaussienne donnée par

$$C_{ga}(u, v) = \Phi_2(\Phi^{-1}(u), \Phi^{-1}(v)); \rho),$$

où Φ est la fonction de répartition unidimensionnelle de la loi normale centrée réduite et Φ_2 est la fonction de répartition bidimensionnelle d'une loi normale centrée de coefficient de corrélation ρ . On voit qu'on obtient l'indépendance lorsque $\rho = 0$.

Soit $V_1 \sim N(\mu, \sigma)$ et $V_2 \sim \exp(\lambda)$. Pour les simulations, on choisit $N = 1000$ et on génère l'échantillon $\{(Y_{1i}, Y_{2i}, X_i)\}_{i=1}^n$ selon les deux modèles suivants :

DGP1

1. $X \sim U[-1, 1]$,
2. $(U_1, U_2) \sim C_{gu}(\cdot, \cdot | \theta = X^2 + 1)$,
3. $Y_1 = F_{V_1}^{-1}(U_1, \mu = X, \sigma = 1)$ et $Y_2 = F_{V_2}^{-1}(U_2, \lambda = 1)$,

DGP2

1. $X \sim N(0, 1)$,
2. $(U_1, U_2) \sim C_{ga}(\cdot, \cdot | \theta = X)$,
3. $Y_1 = F_{V_1}^{-1}(U_1, \mu = X + 1, \sigma = 1)$ et $Y_2 = F_{V_1}^{-1}(U_2, \mu = 0, \sigma = 1)$,

On applique le test pour les deux DGP et on prend soin de changer la valeur de la covariable selon H_0 ou H_1 . On retrouve les résultats pour le niveau du test dans le Tableau 3.1. La puissance du test est illustrée dans le Tableau 3.2.

TABLE 3.1 – Niveau du test d'indépendance conditionnelle calculé avec $x = 0$ pour le DGP1 et $x = 0$ pour le DGP2.

	DGP1	DGP2	DGP1	DGP2
	$\alpha = 5\%$		$\alpha = 10\%$	
$n = 250$	0.064	0.059	0.135	0.11
$n = 500$	0.071	0.053	0.114	0.121

Selon le Tableau 3.1, le test semble mieux performer pour le deuxième modèle. Le niveau est un peu au dessus de ce qu'il devrait être pour le premier test (6.4% vs 5% et 7.1% vs 5%). La taille d'échantillon semble avoir un meilleur impact sur le deuxième modèle.

Le test n'est pas parfait et on peut sûrement améliorer son efficacité en incluant des méthodes de sélection du paramètre de lissage. Regardons maintenant le Tableau 3.2. Le test fonctionne bien dans les deux modèles. On note une augmentation de la puissance lorsque la taille d'échantillon augmente.

TABLE 3.2 – Puissance du test d'indépendance conditionnelle calculée avec $x = 3$ pour le DGP1 et $x = 0.5$ pour le DGP2.

	DGP1	DGP2	DGP1	DGP2
	$\alpha = 5\%$		$\alpha = 10\%$	
$n = 250$	0.998	0.961	1	1
$n = 500$	1	1	1	1

CONCLUSION

Dans ce mémoire, nous avons développé un test d'indépendance conditionnelle basé sur un processus de copule conditionnelle. Ce processus a été proposé et étudié dans [VOG11]. Plus précisément, nous avons développé un estimateur de la dérivée de la fonction de répartition conditionnelle. La convergence en loi ainsi que la convergence uniforme en probabilité de cet estimateur ont été démontrées. En utilisant cet estimateur, on a estimé le biais asymptotique de H_x sous l'hypothèse nulle. L'algorithme nous permettant de répliquer le processus de copule conditionnelle a été établi pour obtenir un échantillon de la statistique de Cramér-von Mises. Pour finir, on a étudié le niveau et la puissance du test à travers deux modèles utilisés pour les simulations. Le choix du paramètre de lissage reste encore à être étudié. On peut même penser au choix de h_n pour l'estimation de \dot{F}_{xh} . On peut aussi baser le test sur une technique de rééchantillonnage utilisée dans [BRT12].

ANNEXE

Démonstration (Proposition 1.1). Pour les démonstrations, on ajoute et on soustrait l'espérance de l'expression. On utilise ensuite la définition de la convergence en probabilité et l'inégalité de Tchebychev.

$$\begin{aligned}
 1. \quad \mathbb{E} \left[K \left(\frac{X_i - x}{h_n} \right) \right] &= \int_{x-h_n}^{x+h_n} K \left(\frac{z-x}{h_n} \right) \mu(z) dz \\
 &= h_n \int_{-1}^1 K(v) \mu(x + h_n v) dv \\
 &= h_n \left\{ \int_{-1}^1 K(v) [\mu(x) + \mu'(x) h_n v + \frac{1}{2} \mu''(x) h_n^2 v^2] dv + o(h_n^2) \right\} \\
 &= h_n \mu(x) + h_n^3 \frac{1}{2} \mu''(x) + o(h_n^3)
 \end{aligned} \tag{3.1}$$

$$\begin{aligned}
 \mathbb{E} \left[\left\{ K \left(\frac{X_i - x}{h_n} \right) \right\}^2 \right] &= \int_{x-h_n}^{x+h_n} K^2 \left(\frac{z-x}{h_n} \right) \mu(z) dz \\
 &= h_n \int_{-1}^1 K^2(v) \mu(x + h_n v) dv \\
 &= h_n \int_{-1}^1 K^2(v) [\mu(x) + o(1)] dv \\
 &= h_n R(K) \mu(x) + o(h_n).
 \end{aligned}$$

$$\begin{aligned}
\text{Var} \left(\frac{1}{nh_n} \sum_{i=1}^n K \left(\frac{X_i - x}{h_n} \right) \right) &= \frac{1}{nh_n^2} \text{Var} \left(K \left(\frac{X_i - x}{h_n} \right) \right) \\
&= \frac{1}{nh_n} R(K) \mu(x) - \frac{1}{n} \mu^2(x) + o(n^{-1}).
\end{aligned} \tag{3.2}$$

On considère ensuite l'égalité suivante :

$$\hat{\mu}(x) - \mu(x) = \hat{\mu}(x) - \mathbb{E}(\hat{\mu}(x)) + \mathbb{E}(\hat{\mu}(x)) - \mu(x).$$

On montre que $\hat{\mu}(x) = \frac{1}{nh_n} \sum_{i=1}^n K \left(\frac{X_i - x}{h_n} \right)$ converge en probabilité vers $\mathbb{E}(\hat{\mu}(x))$ par l'inégalité de Tchebychev et du fait que la variance de $\hat{\mu}(x)$ est un terme négligeable qui tend vers 0. Sous les conditions sur le paramètre de lissage et selon les équations (3.1) et (3.2), on en déduit que

$$\hat{\mu}(x) \xrightarrow{P} \mu(x).$$

Pour les prochaines démonstrations de cette proposition, il suffit de calculer l'espérance et la variance de chaque expression et de procéder avec le même raisonnement.

$$\begin{aligned}
2. \quad \mathbb{E} \left[K \left(\frac{X_i - x}{h_n} \right) (X_i - x) \right] &= h_n^2 \int_{-1}^1 K(v) v \mu(x + h_n v) dv \\
&= h_n^2 \int_{-1}^1 K(v) v [\mu(x) + \mu'(x) h_n v + o(h_n)] dv \\
&= h_n^2 \mu(x) \int_{-1}^1 K(v) v dv + h_n^3 \kappa \mu'(x) + o(h_n^3) \\
&= h_n^3 \kappa \mu'(x) + o(h_n^3).
\end{aligned} \tag{3.3}$$

$$\begin{aligned}
\mathbb{E} \left[\left\{ K \left(\frac{X_i - x}{h_n} \right) (X_i - x) \right\}^2 \right] &= h_n^3 \int_{-1}^1 K^2(v) v^2 \mu(x + h_n v) dv \\
&= h_n^3 \int_{-1}^1 K^2(v) v^2 [\mu(x) + o(1)] dv \\
&= h_n^3 D_1 \mu(x) + o(h_n^3) \quad \text{où } D_1 \in \mathbb{R}^*.
\end{aligned}$$

$$\begin{aligned}
\text{Var} \left(\frac{1}{nh_n^3} \sum_{i=1}^n K \left(\frac{X_i - x}{h_n} \right) (X_i - x) \right) &= \frac{1}{nh_n^6} \text{Var} \left(K \left(\frac{X_i - x}{h_n} \right) (X_i - x) \right) \\
&= \frac{1}{nh_n^3} D_1 \mu(x) - \frac{1}{n} \kappa^2 (\mu'(x))^2 + o(n^{-1}).
\end{aligned} \tag{3.4}$$

Selon les équations (3.3) et (3.4) ainsi que les conditions sur le paramètre de lissage h_n , on obtient :

$$\Rightarrow \frac{1}{nh_n^3} \sum_{i=1}^n K \left(\frac{X_i - x}{h_n} \right) (X_i - x) \xrightarrow{P} \kappa \mu'(x).$$

$$\begin{aligned}
3. \quad \mathbb{E} \left[K \left(\frac{X_i - x}{h_n} \right) (X_i - x)^2 \right] &= h_n^3 \int_{-1}^1 K(v) v^2 \mu(x + h_n v) dv \\
&= h_n^3 \int_{-1}^1 K(v) v^2 [\mu(x) + o(1)] dv \\
&= h_n^3 \mu(x) \int_{-1}^1 K(v) v^2 dv + o(h_n^3) \\
&= h_n^3 \kappa \mu(x) + o(h_n^3).
\end{aligned} \tag{3.5}$$

$$\begin{aligned}
\mathbb{E} \left[\left\{ K \left(\frac{X_i - x}{h_n} \right) (X_i - x)^2 \right\}^2 \right] &= h_n^5 \int_{-1}^1 K^2(v) v^4 \mu(x + h_n v) dv \\
&= h_n^5 \int_{-1}^1 K^2(v) v^4 [\mu(x) + o(1)] dv \\
&= h_n^5 D_2 \mu(x) + o(h_n^5) \quad \text{où } D_2 \in \mathbb{R}^*.
\end{aligned}$$

$$\begin{aligned}
\text{Var} \left(\frac{1}{nh_n^3} \sum_{i=1}^n K \left(\frac{X_i - x}{h_n} \right) (X_i - x)^2 \right) &= \frac{1}{nh_n^6} \text{Var} \left(K \left(\frac{X_i - x}{h_n} \right) (X_i - x)^2 \right) \\
&= \frac{1}{nh_n} D_2 \mu(x) - \frac{1}{n} \kappa^2 \mu^2(x) + o(n^{-1}).
\end{aligned} \tag{3.6}$$

Selon les équations (3.5) et (3.6) ainsi que les conditions sur le paramètre de lissage h_n , on obtient :

$$\Rightarrow \frac{1}{nh_n^3} \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right) (X_i - x)^2 \xrightarrow{P} \kappa\mu(x).$$

4. On a déjà que

$$\mathbb{E} \left[\left\{ K\left(\frac{X_i - x}{h_n}\right) \right\}^2 \right] = h_n R(K)\mu(x) + o(h_n^2). \quad (3.7)$$

$$\begin{aligned} \mathbb{E} \left[\left\{ K\left(\frac{X_i - x}{h_n}\right) \right\}^4 \right] &= h_n \int_{-1}^1 K^4(v) \mu(x + h_n v) dv \\ &= h_n \int_{-1}^1 K^4(v) [\mu(x) + o(1)] dv \\ &= h_n D_3 \mu(x) + o(h_n) \quad \text{où } D_3 \in \mathbb{R}^*. \end{aligned}$$

$$\begin{aligned} \text{Var} \left(\frac{1}{nh_n} \sum_{i=1}^n \left\{ K\left(\frac{X_i - x}{h_n}\right) \right\}^2 \right) &= \frac{1}{nh_n^2} \text{Var} \left(\left\{ K\left(\frac{X_i - x}{h_n}\right) \right\}^2 \right) \\ &= \frac{1}{nh_n} D_3 \mu(x) - \frac{1}{n} (R(K))^2 \mu^2(x) + o(n^{-1}). \end{aligned} \quad (3.8)$$

Selon les équations (3.7) et (3.8) ainsi que les conditions sur le paramètre de lissage h_n , on obtient :

$$\Rightarrow \frac{1}{nh_n} \sum_{i=1}^n \left\{ K\left(\frac{X_i - x}{h_n}\right) \right\}^2 \xrightarrow{P} R(K)\mu(x).$$

Démonstration (Proposition 1.2).

$$\begin{aligned}
\mathbb{E} \left(\mathbb{1}(Y \leq y) K \left(\frac{X - x}{h_n} \right) \right) &= \mathbb{E} \left(K \left(\frac{X - x}{h_n} \right) \mathbb{E}(\mathbb{1}(Y \leq y) | X) \right) \\
&= \mathbb{E} \left(K \left(\frac{X - x}{h_n} \right) F_X(y) \right) \\
&= \int_{x-h_n}^{x+h_n} K \left(\frac{z - x}{h_n} \right) F_z(y) \mu(z) dz \\
&= \int_{-1}^1 K(v) F_{x+h_nv}(y) \mu(x + h_nv) h_n dv \\
&= h_n \int_{-1}^1 K(v) F_{x+h_nv}(y) \mu(x + h_nv) dv \\
&= h_n \int_{-1}^1 K(v) \left[F_x(y) + \dot{F}_x(y) h_nv + \frac{1}{2} \ddot{F}_x(y) h_n^2 v^2 \right] \\
&\quad \cdot \left[\mu(x) + \dot{\mu}(x) h_nv + \frac{1}{2} \ddot{\mu}(x) h_n^2 v^2 \right] dv + o(h_n^2) \\
&= h_n F_x(y) \mu(x) + \frac{h_n^3 \kappa}{2} \left[\mu(x) \ddot{F}_x(y) + \right. \\
&\quad \left. + 2 \dot{\mu}(x) \dot{F}_x(y) + \ddot{\mu} F_x(y) \right] + o(h_n^3).
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}(F_{xh}(y) \hat{\mu}(x)) &= \mathbb{E} \left(\frac{1}{nh_n} \frac{\sum_{i=1}^n \mathbb{1}(Y_i \leq y) K \left(\frac{X_i - x}{h_n} \right)}{\hat{\mu}(x)} \hat{\mu}(x) \right) \\
&= \mathbb{E} \left(\frac{1}{nh_n} \sum_{i=1}^n \mathbb{1}(Y_i \leq y) K \left(\frac{X_i - x}{h_n} \right) \right) \\
&= \frac{1}{h_n} \mathbb{E} \left(\mathbb{1}(Y \leq y) K \left(\frac{X - x}{h_n} \right) \right) \\
&= F_x(y) \mu(x) + \frac{h_n^2 \kappa}{2} \left[\mu(x) \ddot{F}_x(y) + 2 \dot{\mu}(x) \dot{F}_x(y) + \ddot{\mu} F_x(y) \right] + o(h_n^2).
\end{aligned}$$

□

On trouve l'espérance de $\tilde{M}_x(y)$.

$$\begin{aligned}
\mathbb{E}(\tilde{M}_x(y)) &= \mathbb{E}(F_{xh}(y) \hat{\mu}(x)) - F_x(y) \mathbb{E}(\hat{\mu}(x)) \\
&= \frac{1}{2} h_n^2 \kappa [\mu(x) \ddot{F}_x(y) + 2 \dot{\mu}(x) \dot{F}_x(y)] + o(h_n^2).
\end{aligned}$$

Démonstration (Proposition 2.1).

$$\begin{aligned}
1. \quad \mathbb{E} \left[\dot{K} \left(\frac{X_i - x}{h_n} \right) \right] &= h_n \int_{-1}^1 \dot{K}(v) \mu(x + h_n v) dv \\
&= h_n \int_{-1}^1 \dot{K}(v) [\mu(x) + \mu'(x) h_n v + o(h_n)] dv \\
&= h_n \int_{-1}^1 \dot{K}(v) dv \mu(x) + h_n^2 \int_{-1}^1 \dot{K}(v) v dv + o(h_n^2) \\
&= h_n (K(1) - K(-1)) \mu(x) \\
&\quad + h_n^2 \left\{ K(v) v \Big|_{-1}^1 - \int_{-1}^1 K(v) dv \right\} + o(h_n^2) \\
&= h_n^2 (2K(1) - 1) \mu'(x) + o(h_n^2). \tag{3.9}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E} \left[\left\{ \dot{K} \left(\frac{X_i - x}{h_n} \right) \right\}^2 \right] &= h_n \int_{-1}^1 \{\dot{K}(v)\}^2 \mu(h_n v + x) dv \\
&= h_n \int_{-1}^1 \{\dot{K}(v)\}^2 [\mu(x) + o(1)] dv \\
&= h_n R(\dot{K}) \mu(x) + o(h_n).
\end{aligned}$$

$$\begin{aligned}
\text{Var} \left(\frac{1}{nh_n^2} \sum_{i=1}^n \dot{K} \left(\frac{X_i - x}{h_n} \right) \right) &= \frac{1}{nh_n^4} \text{Var} \left(\dot{K} \left(\frac{X_i - x}{h_n} \right) \right) \\
&= \frac{1}{nh_n^3} R(\dot{K}) \mu(x) - \frac{1}{n} (2K(1) - 1)^2 (\mu'(x))^2 + o(n^{-1}). \tag{3.10}
\end{aligned}$$

Selon les équations (3.9) et (3.10) ainsi que les conditions sur le paramètre de lissage h_n , on obtient :

$$\Rightarrow \frac{1}{nh_n^2} \dot{K} \left(\frac{X_i - x}{h_n} \right) \xrightarrow{P} (2K(1) - 1) \mu'(x).$$

$$\begin{aligned}
2. \quad \mathbb{E} \left[K \left(\frac{X_i - x}{h_n} \right) \dot{K} \left(\frac{X_i - x}{h_n} \right) (X_i - x) \right] &= h_n^2 \int_{-1}^1 K(v) \dot{K}(v) v \mu(h_n v + x) dv \\
&= h_n^2 \int_{-1}^1 K(v) \dot{K}(v) v [\mu(x) + o(1)] dv \\
&= h_n^2 \mu(x) \int_{-1}^1 K(v) \dot{K}(v) v dv + o(h_n^2) \\
&= h_n^2 \mu(x) I + o(h_n^2). \tag{3.11}
\end{aligned}$$

Or,

$$\begin{aligned}
I &= \int_{-1}^1 K(v) \dot{K}(v) v dv \\
&= K(v) K(v) v \Big|_{-1}^1 - \int_{-1}^1 K(v) [\dot{K}(v) v + K(v)] dv \\
&= 2K^2(1) - I - \int_{-1}^1 K^2(v) dv
\end{aligned}$$

Donc,

$$\int_{-1}^1 K(v) \dot{K}(v) v dv = \frac{1}{2} (2K^2(1) - R(K)).$$

D'un autre côté, on a

$$\begin{aligned}
\mathbb{E} \left[\left\{ K \left(\frac{X_i - x}{h_n} \right) \dot{K} \left(\frac{X_i - x}{h_n} \right) (X_i - x) \right\}^2 \right] &= h_n^3 \int_{-1}^1 K^2(v) \{\dot{K}(v)\}^2 \mu(h_n v + x) dv \\
&= h_n^3 \int_{-1}^1 K^2(v) \{\dot{K}(v)\}^2 [\mu(x) + o(1)] dv \\
&= h_n^3 D_4 \mu(x) + o(h_n^3) \quad \text{où } D_4 \in \mathbb{R}^*.
\end{aligned}$$

Soit $\mathcal{A} = \frac{1}{nh_n^2} \sum_{i=1}^n K \left(\frac{X_i - x}{h_n} \right) \dot{K} \left(\frac{X_i - x}{h_n} \right) (X_i - x)$, alors

$$\begin{aligned}
\text{Var}(\mathcal{A}) &= \frac{1}{nh_n^4} \text{Var} \left(K \left(\frac{X_i - x}{h_n} \right) \dot{K} \left(\frac{X_i - x}{h_n} \right) (X_i - x) \right) \\
&= \frac{1}{nh_n} D_4 \mu(x) - \frac{1}{4n} \mu^2(x) (2K^2(1) - R(K))^2 + o(n^{-1}). \tag{3.12}
\end{aligned}$$

Selon les équations (3.11) et (3.12) ainsi que les conditions sur le paramètre de lissage h_n , on obtient :

$$\frac{1}{nh_n^2} \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right) \dot{K}\left(\frac{X_i - x}{h_n}\right) (X_i - x) \xrightarrow{P} \frac{1}{2}(2K^2(1) - R(K))\mu'(x).$$

$$\begin{aligned} 3. \quad \mathbb{E} \left[\dot{K}\left(\frac{X_i - x}{h_n}\right) (X_i - x) \right] &= h_n^2 \int_{-1}^1 \dot{K}(v) v \mu(x + h_n v) dv \\ &= h_n^2 \int_{-1}^1 \dot{K}(v) v [\mu(x) + o(1)] dv \\ &= h_n^2 \mu(x) \int_{-1}^1 \dot{K}(v) v dv + o(h_n^2) \\ &= h_n^2 \mu(x) \left\{ K(v) v \Big|_{-1}^1 - \int_{-1}^1 K(v) dv \right\} + o(h_n^2) \\ &= h_n^2 (2K(1) - 1) \mu(x) + o(h_n^2). \end{aligned} \quad (3.13)$$

$$\begin{aligned} \mathbb{E} \left[\left\{ \dot{K}\left(\frac{X_i - x}{h_n}\right) (X_i - x) \right\}^2 \right] &= h_n^3 \int_{-1}^1 \{\dot{K}(v)\}^2 v^2 \mu(h_n v + x) dv \\ &= h_n^3 D_5 \mu(x) + o(h_n^3) \quad \text{où } D_5 \in \mathbb{R}^*. \end{aligned}$$

$$\begin{aligned} \text{Var} \left(\frac{1}{nh_n^2} \sum_{i=1}^n \dot{K}\left(\frac{X_i - x}{h_n}\right) (X_i - x) \right) &= \frac{1}{nh_n^4} \text{Var} \left(\dot{K}\left(\frac{X_i - x}{h_n}\right) (X_i - x) \right) \\ &= \frac{1}{nh_n} D_5 \mu(x) - \frac{1}{n} (2K(1) - 3\kappa)^2 (\mu'(x))^2 + o(n^{-1}). \end{aligned} \quad (3.14)$$

Selon les équations (3.13) et (3.14) ainsi que les conditions sur le paramètre de lissage h_n , on obtient :

$$\Rightarrow \frac{1}{nh_n^2} \sum_{i=1}^n \dot{K}\left(\frac{X_i - x}{h_n}\right) (X_i - x) \xrightarrow{P} (2K(1) - 1) \mu(x).$$

$$\begin{aligned}
4. \quad \mathbb{E} \left[\dot{K} \left(\frac{X_i - x}{h_n} \right) (X_i - x)^2 \right] &= h_n^3 \int_{-1}^1 \dot{K}(v) v^2 \mu(x + h_n v) dv \\
&= h_n^3 \int_{-1}^1 \dot{K}(v) v^2 [\mu(x) + \mu'(x) h_n v + o(h_n)] dv \\
&= h_n^3 \int_{-1}^1 \dot{K}(v) v^2 dv + h_n^4 \int_{-1}^1 \dot{K}(v) v^3 dv \mu'(x) + o(h_n^3) \\
&= h_n^3 \left\{ K(v) v^2 \Big|_{-1}^1 - 2 \int_{-1}^1 K(v) v dv \right\} + h_n^4 \left\{ K(v) v^3 \Big|_{-1}^1 \right. \\
&\quad \left. - 3 \int_{-1}^1 K(v) v^2 dv \right\} \mu'(x) + o(h_n^4) \\
&= h_n^4 (2K(1) - 3\kappa) \mu'(x) + o(h_n^4). \tag{3.15}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E} \left[\left\{ \dot{K} \left(\frac{X_i - x}{h_n} \right) (X_i - x)^2 \right\}^2 \right] &= h_n^5 \int_{-1}^1 \{\dot{K}(v)\}^2 v^4 \mu(h_n v + x) dv \\
&= h_n^5 \int_{-1}^1 \{\dot{K}(v)\}^2 v^4 [\mu(x) + o(1)] dv \\
&= h_n^5 D_6 \mu(x) + o(h_n^5) \quad \text{où } D_6 \in \mathbb{R}^*.
\end{aligned}$$

$$\begin{aligned}
\text{Var} \left(\frac{1}{nh_n^4} \sum_{i=1}^n \dot{K} \left(\frac{X_i - x}{h_n} \right) (X_i - x)^2 \right) &= \frac{1}{nh_n^8} \text{Var} \left(\dot{K} \left(\frac{X_i - x}{h_n} \right) (X_i - x)^2 \right) \\
&= \frac{1}{nh_n^3} D_6 \mu(x) - \frac{1}{n} (2K(1) - 3\kappa)^2 (\mu'(x))^2 + o(n^{-1}). \tag{3.16}
\end{aligned}$$

Selon les équations (3.15) et (3.16) ainsi que les conditions sur le paramètre de lissage h_n , on obtient :

$$\Rightarrow \frac{1}{nh_n^4} \sum_{i=1}^n \dot{K} \left(\frac{X_i - x}{h_n} \right) (X_i - x)^2 \xrightarrow{P} (2K(1) - 3\kappa) \mu'(x).$$

5. On a déjà que

$$\mathbb{E} \left[\left\{ \dot{K} \left(\frac{X_i - x}{h_n} \right) \right\}^2 \right] = h_n R(\dot{K}) \mu(x) + o(h_n). \tag{3.17}$$

$$\begin{aligned}
\mathbb{E} \left[\left\{ \dot{K} \left(\frac{X_i - x}{h_n} \right) \right\}^4 \right] &= h_n \int_{-1}^1 \{ \dot{K}(v) \}^4 \mu(h_n v + x) dv \\
&= h_n \int_{-1}^1 \{ \dot{K}(v) \}^4 [\mu(x) + o(1)] dv \\
&= h_n D_7 \mu(x) + o(h_n) \quad \text{où } D_7 \in \mathbb{R}^*.
\end{aligned}$$

$$\begin{aligned}
\text{Var} \left(\frac{1}{nh_n} \sum_{i=1}^n \left\{ \dot{K} \left(\frac{X_i - x}{h_n} \right) \right\}^2 \right) &= \frac{1}{nh_n^2} \text{Var} \left(\left\{ \dot{K} \left(\frac{X_i - x}{h_n} \right) \right\}^4 \right) \\
&= \frac{1}{nh_n} D_7 \mu(x) - \frac{1}{n} (R(\dot{K})^2 \mu^2(x) + o(n^{-1})).
\end{aligned} \tag{3.18}$$

Selon les équations (3.17) et (3.18) ainsi que les conditions sur le paramètre de lissage h_n , on obtient :

$$\Rightarrow \frac{1}{nh_n} \sum_{i=1}^n \left\{ \dot{K} \left(\frac{X_i - x}{h_n} \right) \right\}^2 \xrightarrow{P} R(\dot{K}) \mu(x).$$

$$\begin{aligned}
6. \quad \mathbb{E} \left[\left\{ \dot{K} \left(\frac{X_i - x}{h_n} \right) \right\}^2 (X_i - x) \right] &= h_n^2 \int_{-1}^1 \{ \dot{K}(u) \}^2 v \mu(h_n v + x) dv \\
&= h_n^2 \int_{-1}^1 \{ \dot{K}(u) \}^2 v [\mu(x) + \mu'(x) h_n v + o(h_n)] dv \\
&= \mu(x) \int_{-1}^1 \{ \dot{K}(u) \}^2 v dv h_n^2 + h_n^3 \mu'(x) \int_{-1}^1 \{ \dot{K}(u) \}^2 v^2 dv + o(h_n^3).
\end{aligned}$$

On utilise le fait que \dot{K} est une fonction anti-symétrique et on a que

$$\begin{aligned}
\int_{-1}^1 (\dot{K}(u))^2 u du &= \int_{-1}^0 (\dot{K}(u))^2 u du + \int_0^1 (\dot{K}(u))^2 u du \\
&= \int_{-1}^0 (\dot{K}(u))^2 u du + \int_0^{-1} (\dot{K}(-v))^2 (-v)(-dv), \quad v = -u \\
&= \int_{-1}^0 (\dot{K}(u))^2 u du + \int_0^{-1} (-\dot{K}(v))^2 v dv \\
&= \int_{-1}^0 (\dot{K}(u))^2 u du - \int_{-1}^0 (\dot{K}(v))^2 v dv \\
&= 0
\end{aligned}$$

Le premier terme du développement de $E \left[\left\{ \dot{K} \left(\frac{X_i - x}{h_n} \right) \right\}^2 (X_i - x) \right]$ s'annule et on a :

$$E \left[\left\{ \dot{K} \left(\frac{X_i - x}{h_n} \right) \right\}^2 (X_i - x) \right] = D_8 \mu'(x) h_n^3 + o(h_n^3) \quad \text{où } D_8 \in \mathbb{R}^*. \quad (3.19)$$

$$\begin{aligned}
E \left[\left\{ \dot{K} \left(\frac{X_i - x}{h_n} \right) \right\}^4 (X_i - x)^2 \right] &= h_n^3 \int_{-1}^1 \{ \dot{K}(v) \}^4 v^2 \mu(h_n v + x) dv \\
&= h_n^3 \int_{-1}^1 \{ \dot{K}(v) \}^4 v^2 [\mu(x) + o(1)] dv \\
&= h_n^3 D_9 \mu(x) + o(h_n^3) \quad \text{où } D_9 \in \mathbb{R}^*.
\end{aligned}$$

$$\begin{aligned}
\text{Var} \left(\frac{1}{n h_n^3} \sum_{i=1}^n \left\{ \dot{K} \left(\frac{X_i - x}{h_n} \right) \right\}^2 (X_i - x) \right) &= \frac{1}{n h_n^6} \text{Var} \left(\left\{ \dot{K} \left(\frac{X_i - x}{h_n} \right) \right\}^2 (X_i - x) \right) \\
&= \frac{1}{n h_n^3} D_9 \mu(x) - \frac{1}{n} D_8^2 (\mu'(x))^2 + o(n^{-1}).
\end{aligned} \quad (3.20)$$

Selon les équations (3.19) et (3.20) ainsi que les conditions sur le paramètre de lissage h_n , on obtient :

$$\Rightarrow \frac{1}{n h_n^3} \sum_{i=1}^n \left\{ \dot{K} \left(\frac{X_i - x}{h_n} \right) \right\}^2 (X_i - x) \xrightarrow{P} D_8 \mu'(x).$$

Bibliographie

- [BHP98] A. Bowman, P. Hall, and T. Prvanm. Bandwidth selection for the smoothing of distribution functions. *Biometrika*, 85 :799–808, 1998.
- [BRT12] T. Bouezmarni, J. V.K. Rombouts, and A. Taamouti. Nonparametric copula-based test for conditional independence with applications to granger causality. *Journal of Business and Economic Statistics*, 30 :275–287, 2012.
- [CB01] G. Casella and R. L. Berger. *Statistical Inference*. Duxbury Press, Duxbury, 2001.
- [Deh79] P. Deheuvels. La fonction de dépendance empirique et ses propriétés : Un test nonparamétrique d’indépendance. *Bulletin de l’académie Royal de Belgique*, pages 274–292, 1979.
- [Deh80] P. Deheuvels. Non parametric tests of independence. *Nonparametric Asymptotic Statistics*, 821 :95–107, 1980.
- [Deh81] P. Deheuvels. A kolmogorov–smirnov type test for independence in multivariate samples. *Rev. Roum. Math. Pures Appl*, 26 :213–226, 1981.
- [Dut15] S. Dutta. Local smoothing for kernel distribution function estimation. *Communications in Statistics : Simulation and Computation*, 44 :878–891, 2015.
- [GF96] I. Gijbels and J. Fan. *Local Polynomial Modelling and Its Applications*. Chapman and Hall, Londres, 1996.

- [GR04] C. Genest and B. Rémillard. Test of independence and randomness based on the empirical copula process. *Test*, 13 :335–369, 2004.
- [GVO10] I. Gijbels, N. Veraverbeke, and M. Omelka. Conditionnal copulas, association measures and their applications. *Computational Statistics and Data Analysis*, 55 :1919–1932, 2010.
- [Ize91] A. J. Izenman. Developments in nonparametric density estimation. *Journal of the American Statistical Association*, 86 :205–224, 1991.
- [Joe14] H. Joe. *Dependence modelling with Copulas*. Chapman and Hall, 2014.
- [JW94] M.C Jones and M. Wand. *Kernel smoothing*. Chapman and Hall, 1994.
- [LLR13] Q. Li, J. Lin, and J. Racine. Optimal bandwidth selection for nonparametric conditional distribution and quantile functions. *Journal of Business and Economic Statistics*, 31 :57–65, 2013.
- [Loa99] C. Loader. Bandwidth selection : Classical or plug-in ?. *The Annals of Statistics*, 27 :415–438, 1999.
- [MP87] J. S. Marron and W. J. Padgett. Asymptotically optimal bandwidth selection for kernel density estimator for randomly right-censored samples. *The Annals of Statistics*, 15 :1520–1535, 1987.
- [Nad64] E. A. Nadaraya. On estimating regression. *Theory of Probability and its Applications*, 9 :141–2, 1964.
- [Nel07] R. Nelsen. *An introduction to copulas*. Springer, 2007.
- [Pat06] A. J. Patton. Modelling asymmetric exchange rate dependence. *International economic review*, 47, 2006.
- [RL06] J. Racine and Q. Li. *Nonparametric Econometrics : Theory and Practice*. Princeton University Press, 2006.

- [Sar93] P. Sarda. Smoothing parameter selection for smooth distribution functions. *Journal of Statistical Planning and Inference*, 35 :64–75, 1993.
- [She04] S. J. Sheater. Density estimation. *Statistical Science*, 19 :588–597, 2004.
- [Sil86] B. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, 1986.
- [Skl59] M. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8 :229–231, 1959.
- [Stu86] W. Stute. Conditionnal empirical processes. *The Annals of Statistics*, 14 :638–647, 1986.
- [vdV00] A. W. van der Vaart. *Asymptotic statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2000.
- [VGO14] N. Veraverbeke, I. Gijbels, and M. Omelka. Preadjusted non-parametric estimation of a conditional distribution function. *Journal of the Royal Statistical Society*, 76 :399–438, 2014.
- [VOG11] N. Veraverbeke, M. Omelka, and I. Gijbels. Estimation of conditionnal copula and association measures. *Scandinavian Journal of Statistics*, 38 :766–780, 2011.
- [Wat64] G. S. Watson. Smooth regression analysis. *Sankhyā : the Indian Journal of Statistics*, 26 :359–372, 1964.